Genetic Structure of the Maya in Guatemala: Perspectives on the Population History of the Maya using mtDNA and Y-chromosome Markers

By

Anne Elizabeth Justice

Submitted to the graduate degree program in Anthropology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Chairperson: Michael H. Crawford, PhD

_____
James Mielke, PhD

_____
Brent Metz, PhD

_____
Bartholomew Dean, PhD

_____
John Kelly, PhD

_____
Kristin Young, PhD

Date Defended: 11/30 /2011

The Dissertation Committee for Anne Justice
certifies that this is the approved version of the following dissertation:

**Genetic Structure of the Maya in Guatemala:
Perspectives on the Population History of the Maya using mtDNA and Y-chromosome Markers**

_____

Chairperson: Michael H. Crawford, PhD

Date approved: 11/30/11

ABSTRACT

The Maya are a diverse ethno-linguistic group with a rich history and an important historical role in Latin America. While they are often treated as a homogenous group among biologists and physical anthropologists, given their wide geographic occupation, long history, linguistic variability, and cultural variation, there likely exists a quantifiable difference among the Maya linguistic and cultural groups. This dissertation tests the efficacy of treating the Maya populations as a genetically homogenous group genetically distinct from surrounding Meso- and Central American populations. Mitochondrial DNA and Y chromosome data were examined using Mayan populations available in the literature along with two newly sampled populations from Guatemala, the Ch'orti' and Poqomchi'. Poqomchi' Maya is spoken in the eastern Guatemalan highlands. This population is of particular interest as their population's history has resulted in relative isolation from Europeans and non-Mayan populations. Ch'orti' Mayans in eastern Guatemala represent the only likely descendants of the Central Maya region remaining in Guatemala and have resided on the edge of the Maya cultural zone for nearly 2,000 years. Ch'orti' history has likely allowed for a higher degree of non-Mayan and non-native admixture than found among other Mayans including the Poqomchi'.

The lineages present in the Poqomchi' and Ch'orti' are consistent with previous findings in the Mesoamerican cultural and geographical region, but also reflect their unique histories. Despite nearly 500 years of colonization, the Poqomchi' and Ch'orti' Maya maintain a majority of Native American mtDNA (A, B, and C) and Y chromosome (Q) lineages. The Poqomchi' exhibit no maternal non-native gene flow, only one non-native male haplogroup, and have maintained moderate levels of genetic diversity. However, due to their different history during conquest and European commercial industries, the Ch'orti' exhibit a higher degree of admixture for both the paternal and maternal side (26.3% and 5.3% respectively). Also, the Ch'orti' have decreased diversity compared to the Poqomchi' and exhibit structure in their mtDNA lineages in the network analyses. This indicates that they have suffered greater genetic losses and a slower population recovery since colonization and later suffered a slowing of population growth during the Guatemalan Civil War.

This study shows consistently that the Mayan populations share a common history and close genetic relationship, but are not a homogenous population. The mtDNA AMOVA and network analyses for the Maya exhibited little population substructure. This study supports previous findings that the K'iche Maya of the southern Maya area are the most divergent of the Mayan due to their lack of the mtDNA A2 haplotype and the higher frequency of haplogroup D. The results of this study provide evidence of close genetic relationship among Mayan populations to other Mesoamerican populations. The Mesoamerican populations share a common ancestral population and the unique similarities among Mesoamerican populations are maintained through gene flow. In contrast, there is greater genetic structure for both mtDNA and Y chromosome markers across Central and South America. Once again, molecular markers have proven useful in elucidating the historical, geographic, and linguistic relationship among recently diverged human populations.

DEDICATION

I dedicate this work to my family. My husband, Richard, for his continued support and love even across great distances; my children, Liberty, Augustus, and Lily-May, for showing me the patience I don't always show them; my sister, Cathy, for being more than a sister when I needed her most; my mother, Anne, and my father, Joseph, for always encouraging me; and Geetha Chittoor, while not related to me by blood, she is a sister in my heart and has always treated me as family. They have all sacrificed so much to make this possible.

ACKNOWLEDGEMENTS

I would like to thank the indigenous people of Guatemala that participated in and/or supported this study.

I would like to thank the Ministerio de Salud of Guatemala for their support both locally and from the capital. I would especially like to thank Daniel Herrera who helped me maneuver through the process of approval for my research and volunteering in the field. I would also like to give thanks to the other volunteers from the Ministerio de Salud in Tahamú, Herman, Herbert, Ramón, Efrain, and David.

Thank you to Dr. Moses Schanfield at George Washington University; Dr. Pat Williams, Torrey Parrish, and Jennifer Gerst at Evogen for providing me with lab space and training to complete my dissertation labwork; and Dr. Mike Grose at the KU DNA Sequencing Laboratory.

I also give thanks to my committee members, Dr. Michael Crawford, Dr. Bart Dean; Dr. Jim Mielke; Dr. Brent Metz, Dr. Kristin Young and Dr. John Kelly; my friend, colleague, and field partner, Emilia Barbosa; my fellow collaborator on this project, James Herynk; my fellow current and former students from the Laboratory of Biological Anthropology specifically Dr. Geetha Chitoor, Dr. Rohina Rubicz, Dr. Phil Melton, Master Kristie Beaty, Chris Phillips-Krawzack, and Stephen Johnson; the Department of Anthropology office staff: Judy Ross, Carol Archinal, Kathleen Womack, and Le-Thu A. Erazmus; and the faculty of the Department of Anthropology, the Environmental Studies Program, and the Center for Latin American Studies for their continued support throughout the years.

This project was funded by the Latin American Studies Summer Field Research Grant, the Tinker Summer Field Research Grant, and the Carroll D. Clark Award. The project was also funded by the University of Kansas General Research Fund awarded to Dr. Michael Crawford.

TABLE OF CONTENTS

List of Figures

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 3

CHAPTER 4

CHAPTER 1: INTRODUCTION

Recent molecular genetic investigations have shed considerable light on the initial peopling of the Americas. These studies provide evidence for either a single migration or a few migrations from a single source population from Siberia between 18,000 and 21,000 BP (Achilli et al. 2008; de Azevedo et al. 2011; Merriwether et al. 1995; Schurr 2004; Schurr and Sherry 2004; Torroni et al. 1993a). While many of these studies use DNA samples from Maya populations to answer questions regarding the peopling of the Americas as a whole (Achilli et al. 2008; Horai et al. 1993; Schurr et al. 1990), or South America (Lewis et al. 2007; Rothhammer et al. 2001), few studies focus on the population history of the Maya and their influence on migration and gene flow throughout Mesoamerican prehistory or the effects of colonization on their genetic makeup.

In this case, "Maya" refers to a language family, which contains ~30 distinct languages and ethnically diverse populations (Evans and Webster 2001; Ruhlen 1987). However, Maya populations are often lumped together and treated as a homogeneous population, biologically distinct from other Meso- and Central American linguistic and cultural groups. Consequently, there have been conflicting findings regarding the heterogeneity of the Maya and their relationship to surrounding populations. Mesoamerica refers to the distinct cultural and linguistic area, which roughly covers central and southern Mexico, Guatemala, Belize, northern and western Honduras, and El Salvador (Carmack 2007). Mesoamerica is a continuous

geographical area that shares similar cultural traits passed down from the Olmec civilization and evolved together through extensive contact, including shared architecture, maize-based diet, agricultural practices, calendar and cosmological worldview, monetary system, and linguistic traits (Campbell 1997; Carmack et al. 2007). Central America encompasses the isthmus connecting Mesoamerica to South America, Nicaragua, Costa Rica, and Panama. There is a distinct break in cultural traditions and language across this area (Campbell 1997; Carmack et al. 2007; Cooke 2005; Melton 2008). This project will characterize the population structure and make inferences about the prehistory and history of the Maya and surrounding Central Americans by applying multivariate statistics to new and existing molecular data and interpreting the results using previous information regarding geography, archaeology, culture history, and language.

Three major areas (northern, central, and southern) were occupied by the Maya, who can be distinguished on the basis of language, geography, and cultural remains. The Southern Maya Area stretches from the Mexican Chiapas down the western coast of Central America to El Salvador. While this seems to be the first area of settlement for the Maya, it differs archaeologically from Central and Northern Regions. The Central Maya Area covers the entire department of Petén in northern Guatemala, western Honduras, and the lowlands of Belize, northern Chiapas and Tabasco, Mexico. The Central Maya were likely the most diverse and prosperous during the Classic Maya period and represent what most researchers consider, archaeologically "the most typically 'Maya' traits…" (Coe 1966: 27). The Northern Maya Area spread across the arid deserts and coasts of the Yucatan peninsula of Mexico. While archaeologically more similar to the Central Maya than the Southern, the Northern Maya did

not flourish in numbers as did the Central Area during the Classic Period, but they largely displaced the Central Area in importance during the Postclassic Period (Coe 1966). While the Maya likely share a common ancestry with Chibchan-speaking populations to the south, archaeological, linguistic and genetic data shows that there was a barrier between the Maya and their southern neighbors (Campbell 1997; Carmack et al. 2007; Melton 2008). However, some archaeological influences of the Maya can be found among their northern neighbors, the Olmecs and later Aztecs (Coe 2005).

Recent studies which focus on the Maya have used autosomal DNA (Herrera et al. 2007; Ibarra-Rivera et al. 2008), ancient mtDNA (Gonzalez-Oliver et al. 2001) , and odontometrics (Scherer 2007) to shed light on the relationship among members of the Maya language sub-family, but do not include more recently founded populations such as the Poqomchi'-speaking Maya. Studies comparing Mayan populations that are both geographically and/or linguistically isolated (Kaqchikel, K'iche', and Yucatec) have concluded that population sub-structure exists among different Maya-speaking populations. These results indicate that the Maya cannot be considered a homogenous group. These studies also do not speak to their relationship to surrounding Meso- and Central American populations.

However, many modern Maya populations living in Guatemala have no geographic barriers but have a recent linguistic relationship (~1,000 yBP) and populations with a more ancient linguistic relationship (~2-3,000 yBP). Also, no descendants of this Central Maya Area were represented among these samples, and while there are ~30 Mayan languages, only three populations from four geographic locations were used in two of these studies (Herrera et al. 2007; Ibarra-Rivera et al. 2008). The level of similarity among various Maya groups could

change with higher sample numbers and/or more populations sampled. Therefore, studies such as this project are important for understanding the efficacy of viewing the "Maya" as a homogenous group.

This project synthesizes existing biological, archaeological, cultural and linguistic data on the Maya while adding new genetic data (mtDNA and Y-chromosome) from two Maya-speaking populations (Chor'ti' and Poqomchi') to provide insight about the genetic history and diversity of the Maya-speaking peoples. Poqomchi' Maya is a language spoken in the eastern Guatemalan highlands. It is closely related to other core Quichean Mayan languages and Kekchi' Maya. This population is of particular interest for understanding Maya genetic diversity as its history has resulted in relative isolation from European colonists and non-Mayan populations. Additionally, this language has only recently diverged from the other core Quichean languages, making this population ideal for testing hypotheses on the relationship among these language groups (Campbell 1997; Olson 1991; Ruhlen 1987).

Ch'orti' is a language spoken in far eastern Guatemala and western Honduras and El Salvador. This group represents the only surviving descendants of the Central Maya Area in Guatemala, the most prosperous region during the Classic Maya Period (Evans and Webster 2001). In contrast to the Poqomchi', the Ch'orti' have experienced significant admixture with European settlers and Africans brought during the slave trade (Lovell and Lutz 1995). Their close contact and relatively high mobility have also resulted in several population bottlenecks through history. Finally, the Ch'orti' are important to the understanding of Maya phylogeny due to their geographical location on the periphery of the Mayan empire. They offer an excellent opportunity to test hypotheses on the relationship between Maya and non-Maya Central

Americans, as well as contrasting the variable effects of colonization on the Maya. Multivariate analyses will be used to test the following hypotheses:

1) There is significant within-group genetic structure among Maya-speaking populations.

2) The Maya differ genetically from other Meso-, Central, and South American and Caribbean populations.

3) The genetic structure of maternal (inferred from mtDNA) and paternal lineages (inferred from the Y-chromosome) differ from one another.

4) There is a statistical relationship between geography, genetics and languages across Meso-, Central, and South America and the Caribbean.

The chapters that follow will detail the study design, results, and conclusions resulting from this study of Maya biological variation. Chapter two will briefly review the relationship of Mayan languages, the Pre-conquest and Post-conquest history of the Maya geographical region, and the existing literature on the biological makeup of the Mayans. In chapter three, the data set and methods, including the sample collection methods, laboratory methods, and analytical methods will be detailed. Chapter four will present the results of the statistical analyses, which are discussed in context in chapter five. The conclusion will detail Maya history and their relationship to surrounding populations that can be inferred from the data in chapter six.

CHAPTER 2. BACKGROUND AND POPULATION HISTORY

This chapter provides a historical, archaeological, linguistic, and genetic overview of the Poqomchi' and Ch'orti Maya under study, while placing them within the greater context of Meso- and Central America. First, a brief summary of the pre-colonization history is presented focusing on archaeological data that highlight the relationship among Maya populations and surrounding Mesoamericans. This is followed by a discussion of the history of Guatemalan Mayans after colonization. This discussion will highlight the historical events which were most likely to influence the genetic makeup of the population, including migration, population reduction, disease, displacement, and isolation. These sections will also highlight those traits that make Mesoamerica a distinct culture area. Next, background information is provided on the linguistic relationship of the Maya, followed by a review of the current body of literature available about the biology of the Maya and surrounding Native American populations with special attention paid to the molecular markers used in this investigation. Finally, a brief statement is provided on the current history, linguistic, and biological information known about the Poqomchi' and Ch'orti' Mayans.

2.1 PRE-CONQUEST HISTORY

The history of the Maya prior to European contact is divided into three main time periods, the Preclassic (~4000 BP – 1700 BP), the Classic (1700 BP – 1200 BP), and the

Postclassic (1100 BP – shortly after European contact), which can be delineated archaeologically by differences in technology, architecture, art, and military and political organization (Carmack et al. 2007; Coe 2005; Sabloff 1990). Most early archaeologists focused on the Classic period with its grand temples, detailed carvings, and vast geographical range. However, the interpretation of these periods is limited since much of the cultural data focuses on the elite class, paying little attention to the average experiences of the Maya. While this makes interpretation of the archaeological record challenging, several generalities on Maya population lifestyle and movement can be ascertained (Sabloff 1990). The terminology and the distinction of these three time periods are shared with surrounding Mesoamerican populations with whom the Maya shared extensive cultural contact (Coe 2005).

2.1.1 Preclassic period Maya (4000 BP – 1700 BP)

The Preclassic period in the southern Maya Area is characterized by intensified agriculture and settlement (Coe 2005). Preclassic Maya buildings, sculpture, and cultural and biological remains are particularly difficult to interpret. First, Maya structures are often built right on top of older buildings rather than tearing them down (Stanton and Magnoni 2008). Additionally, some architectural or sculptural traits (traditionally defined as Classic period style), can be found much earlier than the 1700 BP cutoff, placing them in the Mesoamerican Preclassic, or much later than 1200 BP, in the Postclassic (Coe 2005; Sabloff 1990).

While it is difficult to piece together details about the Preclassic period, it has been accepted that hunter-gatherer populations lived in the Lowland areas, along with a few larger settlements (e.g. the Olmec along the Caribbean gulf coast, and early Maya in Mirador) (Sabloff 1990). Archaeological remains of human populations in Mesoamerica date back as early as 20,000 BP, long before the populations would be considered Mayan. These early hunter-gatherer populations used simple, unifacially edged and retouched tools and hunted large game animals. Mesoamerica during the Archaic period provided a diverse diet that was easily exploited by the hunter-gatherers, which allowed their population to flourish. At that start of the Preclassic period there was population growth, a shift to sedentary life, and a reduction in the diversity of food resources. Sedentism was followed by an increase in cultivation and domestication of plants and development of pottery that allowed long-term storage of foods (Evans and Webster 2001; Grove 1981; Stark 1981).

These technological advances set the stage for the development and florescence of the Maya during the Preclassic period, with permanent settlements as early as 3200 BP or as late as 2800 BP. These settlers were likely a mix of Maya and non-Maya speaking individuals from the highlands. Their expansion is marked by the first public buildings, likely religious, appearing 2600 BP and followed by a rise in population density. By 2300 BP, the archaeological record reveals urban centers, trade networks, large temples, and increased mobility. During the Late Preclassic (2300 – 1700 BP) temples exhibit the first “high art” that includes plastered masks on the buildings and complex religious symbols (Sabloff 1990).

Soon after the first hunter-gatherer populations became sedentary in the lowlands, there was a demographic population expansion. The increase in population size in the southern

lowlands was due in part to volcanic eruptions in the northern highlands. These volcanic eruptions caused people to relocate to the lowlands where they may have fought for land, and therefore moved to the urban centers for greater protection and access to goods. The soil at that time was rich and able to support growing populations surrounding large city squares. There was great variation in subsistence over both time and space. When the area was first settled, meat likely played a major role in the diet, as well as maize. Over time, agriculture intensified to feed the growing populations and the environment could no longer support the animal population. The diet in the Classic focused more heavily on maize, beans, and squash. By the Late Preclassic period, there were likely more than 1,000,000 inhabitants in the Southern Lowlands. This prosperity marked the trend of population growth, architectural achievement, and high art that continued in the Classic period (Sabloff 1990).

2.1.2 Classic period Maya (1700 BP – 1200 BP)

The Classic Maya period is best known for the grand temples and palaces, a rapid demographic and geographic expansion in the Central Maya Area, and the development and wide-spread use of the long count calendar and a sophisticated writing system (Coe 2005; Evans and Webster 2001). Beginning in the 1950s, the archaeological study of settlement patterns provided new information on the life patterns of Maya peasantry. A more systematic survey of areas surrounding known temples revealed that there were peasant dwellings within

the city core. The spatial arrangements of the housing allowed the researchers to make inferences about the subsistence patterns including agricultural practices (Sabloff 1990).

Mapping of settlement patterns at Tikal, the largest site in Guatemala, revealed an estimated population of 39,000-49,000 individuals residing in and around this city center during the Classic period. The population density was estimated at 600 individuals per square kilometer, indicating that sophisticated agricultural practices must have been used to support such a dense population (Sabloff 1990). Similar settlement patterns were found in Kaminaljuyu in the southern highlands and in Copán in the central Maya Area (Evans and Webster 2001). These findings called into question traditional views, which held that city centers were only occupied by the elite or religious personnel and that the peasants supported these city centers using only slash and burn agriculture. These findings also indicate that the social dynamics of the Maya were more complicated than a simple ruling class and peasant class. Indeed, there is evidence of craft specialization, such as stone workers, farmers, woodworkers, and ceramicists. Specialized agricultural practices included canals, terracing, raised fields, and swamp reclamation (Evans and Webster 2001; Sabloff 1990).

It was not until the late 1950s that major breakthroughs were made in the interpretation of hieroglyphs. These breakthroughs allowed for the recognition of the most important cultural centers for the Classic period, and the kinship of their rulers. Moreover, the realization that these glyphs represent syllables led to an understanding of the ancient phonetics of Mayan languages. With the information contained in these writings, archaeologists now understand more of the Mayan politics, religion, and calendar (Sabloff 1990).

There was great heterogeneity within the Maya civilization of the Classic period both through time and geography. Changes across time include an increased politicization, which resulted in more palaces being built and fewer temples. Also, there was greater concentration of the population in the city centers and greater wealth concentration as well. There were specialized labor centers; for example, stone quarries with tool manufacturing at the Colha site, and areas of intense agriculture and areas with the absence of agriculture (Sabloff 1990).

However, despite their prosperity, the Southern Lowland centers began to decline around 1200 BP. There was a standstill in building; many centers were completely abandoned; and even the less urbanized areas surrounding Maya centers were deserted. While no single answer exists as to why there was a sudden decline in such a successful area, archaeologists have uncovered minor factors that together might explain the end of the Classic Maya Period. Some of the contributing factors include increased social stratification, which limited access to resources for the lower-class leading to malnutrition, increased competition, and increased raids both from within the Maya and from non-Mayan populations. Also, there was a shortage of usable land and viable labor force. The elite seemed to respond to these pressures by increasing agricultural practices, such as swamp reclamation, and increasing monument building to honor the gods. These choices both stressed the land further and pulled workers away from their primary roles in society, weakening the population and making them vulnerable to the environment and warring neighbors (Sabloff 1990).

Simply put, the decline of the Southern Lowlands occurred because the environment could not support an advanced growing population such as the Maya. As agriculture devastated the lands and food became scarce, the Maya had to rely on what little they had to trade. While

their craftsmanship was desired, they had little in the way of raw materials. Those areas that lasted through the Southern decline, were close to water (coastal line, rivers, and lakes) with access to aquatic resources, had some remaining usable soil for growing food, or produced salt (Sabloff 1990). A few coastal southern cities survived the decline, and the Northern Lowlands gained momentum. So, instead of the end of the Maya, this time period is rather a shifting of the Maya (Evans and Webster 2001; Sabloff 1990).

2.1.3 Postclassic period Maya (~1200 BP – 1524 AD)

The shift between the end of the Classic and beginning of the Postclassic is often referred to as the terminal Classic period, which lasted from around 1200 BP to 1000 BP. During this period, there was a shift in geographic distribution, many of the Southern Lowland sites were abandoned and population density shifted toward the Northern Lowland centers. The terminal Classic period was centered on the Puuc Region in the Northern Lowlands. Investigators believe that these cities had stronger economic and social ties than cities in other regions of Maya territory due to extensive road systems connecting the city centers in the Puuc Region. However, this region reached its carrying capacity quickly after its rise, and so too declined by 1000 BP. Toward the end of the Puuc Region rule, Chichén Itzá rose as a cultural, religious and political center which ruled over much of the Northern Lowlands. It remained an important religious center well after the population declined around 800 BP (Sabloff 1990). The

Southern Lowland centers were almost completely abandoned by 900 AD, marking the end of the Classic period.

The Postclassic period is marked by the rapid depopulation of most of the Maya region, but primarily in the Central Area. There is extensive evidence of a cultural invasion of the Aztec (Toltec) Mexican cultures. To many researchers, the Postclassic period marks the end of the Maya. However, some interpret the moving of people to the North and the introduction of Mexican culture into Maya centers as another shift in Maya culture, rather than a replacement (Sabloff 1990).

The Late Postclassic, also referred to as the Decadent Period, lasted from 1250 AD until 1520 AD (Sabloff 1990). The two prominent centers of the northern lowlands for this time period are Mayapán and Tulum, with Mayapán being the larger of the two with an estimated 11,000 inhabitants (Carmack et al. 2007). Both were situated in well protected areas, suggesting that this was a time of frequent wars (Sabloff 1990). The largest remaining Maya center in the Central Area was found at the Taj Itza in the Peten. The southern highlands retained the majority of the operating Mayan states with the largest held by the K'iche Maya at the capital of Utatlan with as many as 15,000 inhabitants. Other highland culture groups with states included the Kaqchikel, Tzutujil, Mam, Ixil, and Poqomam (Carmack et al. 2007). Trade centers were well developed in areas such as Cozumel and Taj Itza, which promoted contact between the three regions and allowed for exchange of regional goods such as quetzal feathers from the highlands and cacao and honey from the lowlands (Carmack et al. 2007; Sabloff 1990). Goods that were previously found only in the residence of the elite were now more commonplace, resulting in the title of the Decadent Period. However, there was still a heavy focus on religion

throughout the Northern Lowlands. Individuals would undertake pilgrimages to older religious sites, such as Chichén Itzá (Sabloff 1990).

As stated above, the Postclassic is marked by a rapid depopulation, especially in the central Maya Area. However, this depopulation did not necessarily result in complete abandonment of sites. In most instances, small enclaves of population remained, temples were used for religious purposes, portable objects were reused, and stone carvings and hieroglyphics were re-carved (Manahan 2008; Stanton and Magnoni 2008). Manahan (2008) argues that this occurred on a large scale after the collapse of the Copán center. The population that moved into these ruins exhibited distinct cultural patterns unlike those present during the Classic Period, indicating that a non-Mayan population, possibly Lenca, resided in Copán at the time of Spanish conquest.

2.1.4 Evidence of Contact between Mayan and Non-Mayan Mesoamericans

Early archaeologists erroneously believed that the Maya civilization was built in isolation from other Meso- and Central American populations, and that there existed little contact among highland and lowland Maya. This is one reason why researchers felt that the Toltec invasion of Chichén Itzá marked the end of the Classic Maya. While there was an increase in the influence of Mexican style after the Toltecs conquered the Maya center, their significant influence could be seen long before this time. Within the Maya there was extensive trade between the lowlands and highlands as early as the late Preclassic. This is evidenced by the

presence of jade, obsidian, and sculptural styles in the lowlands. Furthermore, the Maya hieroglyphics and monumental art likely got their beginnings from the Olmec on the lowland Gulf coast. The presence of Mexican architectural design in Maya centers, such as Tikal, denotes the great influence and extensive trade with the central Mexican center, Teotihuacan. There is evidence that Mexicans from Teotihuacan conquered the Maya center at Kaminaljuyú, just outside modern Guatemala City, in the early Classic period and from there the influence of Mexico spread. In addition to stylistic analyses, the use of chemical analyses helped to date items and locate the source of raw materials. These advanced techniques have allowed archaeologists to demonstrate that the Maya also had an influence over Mexican and Central American culture as well (Cooke 2005; Evans and Webster 2001; Sabloff 1990; Vogt 1969).

The discovery of peripheral fortifications, including ridges, parapets, and ditches; and pictorial and hieroglyphic references to war, indicate that the Maya were not always the peaceful people that traditional archaeologists depicted. Early archaeologists believed that any evidence of warfare was minimal, usually representing small raids between Maya centers. However, new evidence reveals that raids were occurring between Mayan centers as well as between Mexicanized Mayans, Toltecs, and other warriors from central and southern Mexico (Evans and Webster 2001). Two important cities in the lowlands, which alter this view, are Altar de Sacrificios and Seibal, close to the Western edge of the Lowlands in Guatemala where evidence of invasion has been found:

> A wide array of data indicate that non-Classic Maya peoples from the Gulf Coast lowlands had invaded these centers by the beginning of the ninth century AD. These peoples have often been called the Chontal Maya or Putun; they spoke a Maya language but were not part of the Classic Maya civilization. Evidence of the invasion came from ceramic analyses in particular… (Sabloff 1990: 89).

Some indications that these sites were invaded and taken over include changes in the art, hieroglyphs, architecture, and building and center layouts (Coe 2005; Evans and Webster 2001; Sabloff 1990). Hieroglyphs reveal that ties to the Toltec were important for the elite, as Yucatec and Tabasco writings tell of genealogical ties between the Mayan and Toltec rulers. Conversely, K'iche writing tells of military conflict with nearby Mexican states (Carmack et al. 2007; Evans and Webster 2001).

2.2 EUROPEAN CONTACT AND RECENT HISTORY

The conquest of Guatemala took many years, lasting from 1524 AD, when the Spanish first entered the Maya territory, until roughly 1600 AD. Like with the Pre-Contact history, the recent history of Mesoamerica can be divided into three main time periods, primarily determined by major political change. The first is the Colonial Period (~1600 AD – 1821 AD), followed by the Neocolonial Period (1821 AD - 1944 AD), and then by the Modern Period (1944 AD – present).

2.2.1 Mesoamerica at Contact

At the time of Spanish contact, the principal building blocks of Mesoamerica were large towns and their dependent rural communities. Rural communities were made up of kinship

groups (primarily patrilocal), while mainly the elite Mesoamericans lived in the ruling town centers. Although we use the terms empires, states, and chiefdoms to get a broad understanding of the political structures in the area, there was a continuum that ran between chiefdoms to empires. Generally speaking, chiefdoms are based on kinship and allow for the ranking of tribal groups; states allow for social classes of people rather than ranking based solely on kinship (although this was used to determine the elites); and empires consisted of states that ruled over other states (Carmack et al. 2007).

While Mesoamerica was divided into several smaller political units, the broad region can be thought of as a "World System". Mesoamerica was considered a world system, first, because some component societies were dominant over others, creating an integrated, stratified society. Second, no one empire had control; rather there were competing empires and states. Possibly most important, trade held Mesoamerica together under a "world economy." Trade was conducted under two main trade languages, Nahuatl and Chontal Maya (Carmack et al. 2007). There were several centers of political power including, but not limited to, the Aztec empire in Teotihuacán and the northern and southern Maya centers (although these were more a series of connected states rather than empires). The semi-periphery existed on the outskirts of the major centers, and the periphery stretched to the boundaries of Mesoamerica bordering on the wilderness and tribal areas. While there was no free trade in the centers, there were "free-trade centers" in the semi-periphery (much of the trade was still forced trade as any population under rule of a center was required to trade), which were more democratically run by councils and traders. The semi-periphery also acted as religious centers.

The periphery were ruled over and required to provide much of the raw materials traded in the semi-periphery (Carmack et al. 2007).

The Conquest period (1524 AD – 1600 AD) brought a great population decline (estimated as high as 90-95% in many areas) for the Maya and the rest of Meso- and Central America, primarily due to disease, famine, and slavery at the hands of the Europeans (Carmack et al. 2007; Metz 2001). Demographic trends varied greatly from region to region, with the Lowland areas suffering the largest decrease in population size coupled with the slowest recovery (Carmack 1986). After the Spanish came into power over Mesoamerica, the political structure changed as a more European model of political structure was imposed over the natives, with a focus on mining gold. In an attempt to protect the indigenous populations from over exploitation while still using them as a primary labor force, the Spanish crown first implemented the *encomienda*. This was a trustee system of labor set-up by the Spanish in which conquistadors, soldiers and other Spanish representatives were granted a certain number of Indians to direct for the purposes of mining. The *encomenderos* were to teach Spanish and Catholicism to the natives in exchange for tribute, one fifth of which was supposed to be paid in taxes to the Crown. The *encomienda* also consisted of a system of laws meant to end certain abuses against natives, rewarding Spaniards for service to the crown by giving them labor forces if they then paid tribute to the crown (Carmack et al. 2007; Foxen 2007).

2.2.2 Colonial Period

When colonial institutions were established (1600 AD - 1821 AD), a new complicated class system emerged and led to continued political restructuring, and forced migration of native and slave populations around the Caribbean, Meso- and Central America. Under the *encomienda*, the indigenous populations were still overworked and abused, and thus their populations continued to decline, so further changes were made to the political structure. These changes were influenced by new classes and "races." The new colonial caste society was based upon country of origin and racial mixing. The Peninsulares (*Guachupines*), Europeans born in Spain, were the highest ranking class. This class was politically and religiously motivated, but primarily interested in making profit. Creoles were of Spanish descent, but born in the Americas, so had a lower rank then the Peninsulares. Creoles, like their Spanish born counterparts, were interested in holding power and making profit, but found little power or support from the Crown after the end of the *encomienda*. Mestizos, those of mixed indigenous and European descent, were not allowed land, had no special laws protecting them, and could not obtain high roles. They instead took their place as intermediaries in the new social order communicating between the higher social class and the indigenous populations. Mulattos were of mixed African and European descent, found mainly along the coasts where Indians populations had been decimated, and slaves or freed African were brought in or hired as intermediaries or foremen on working fields. Sambos/lobos/Zambos were free slaves and of mixed Native ancestry. Ladino is the term used in Central America (Southern Mexico and Guatemala) to refer to a Christianized non-Spaniard that speaks Spanish. Ladinos had very little status in this social structure and were often displaced workers or moved into cities where they attempted to fit into the new Mesoamerican order. At the bottom of the social class were the

"pure" Indians and African Slaves. There is some argument as to which was considered lower in social status among the two (Carmack et al. 2007; Foxen 2007; Lovell and Lutz 1995). These social and political designations affected the choice of spouse among socio-economic classes. Terminology used to describe the various castes varied among countries and regions; however, the meaning and social implications of these castes were similar throughout Mesoamerica (Gabbert 2004a; Gabbert 2004b).

The primary concern of the Indians was protection of their sacred lands. The Indians often took the upper classes to court in order to protect their home land. While the middle-classes of Creoles and Mestizos had little power or protection under the Crown, they also had little obligation; while the Indians were still required to pay tribute and later to provide forced and often unpaid labor on Spanish farms and in Spanish mines. Many Indians were also forced onto haciendas where they were kept in debt peonage, forced to attend churches, pay tithes, and purchase goods from Spain that they otherwise would not want or need (Carmack et al. 2007; Foxen 2007).

During the Colonial period, closed corporate peasant community models were very important to indigenous populations of Mesoamerica. These communities were united and based highly on communal solidarity and anti-accumulation of wealth. This form of social structure allowed for joint, communal ownership of land, which promoted communication within the community, pooling of resources, and avoidance of envy and competition among the Indians. Therefore, this communal ownership was important in fighting against outside pressures on indigenous communities (Lauria-Santiago and Binford 2004).

2.2.3 Neocolonial Period

Independence from Spain in Meso- and Central America began between 1810 and 1825, led by the creole community who wanted to augment their power (Carmack et al. 2007). Independence from the Spanish for Guatemala and Mexico was achieved in 1821, and the remainder of the neocolonial period saw increased social and geographic mobility of Guatemalan Indians. Also, in the 1880s, there was a movement toward privatization of land in much of Mesoamerica, which interfered with the closed corporate peasant communities. The privatization of land opened up opportunities for corruption within the communities (Coe 2005; Lauria-Santiago and Binford 2004). The corruption and destruction of community led to feelings of distrust within communities and against Ladinos outside of communities. Additionally, the Ladino population began to grow as Indians left their homes to avoid obligation within their communities, a process begun in the Colonial Period. The increase of wealth and the size of the population then led to the rise of a Ladino middle class. Land privatizations, invasions, and renewed forced labor and taxes led to Indian insurrections in late 1800s in places such as El Salvador (Lauria-Santiago and Binford 2004).

The new middle-class was then made up of former displaced Indians and Ladinos who were artisans, technicians, traders, educators, and military officers. The middle class tried to raise themselves up and catch up to the world's middle class. The middle-class was marked by evolutionist thinking and positivist thinking. This class was very nationalistic, anti-communist, and Christian. There was much distrust between the peasant communities and the growing middle-class obsessed with imitating elite life-styles and accumulating wealth (Carmack et al.

2007; Lauria-Santiago and Binford 2004).

2.2.4 Modern Period

As a result of the changing social structure and privatization of land, peasant insurrections began across Mesoamerica in the late 19th century. Many of these revolts resulted in a strong governmental response including mass genocide of indigenous population across El Salvador, Guatemala, and Mexico. The modernization of indigenous peoples of Mesoamerica is viewed by some anthropologists in a positive light, while for others it represents oppression. Where they overlap is in their recognition of the paternalistic treatment of indigenous groups by elites, the state, NGOs; and that this treatment has placed indigenous cultures in a state of crisis. Where they differ is in their attitude about the current result and the future trajectory of the indigenous. One is more positive highlighting new activist movements among the indigenous, advances in organizations and political parties in response to neoliberalism, and a trajectory away from the paternalistic treatment of the state (Dietz 2005; Martínez Novo 2006). Racism continues to occur throughout Mesoamerica, influencing the formulation of identity. Many individuals identify themselves as Ladino denying their biological heritage due to greater access to employment or social capital (Pine 2008).

In the mid-20th century across Mesoamerica rose a movement called *indigenismo*, the idea that the indigenous should be integrated into an idealized mestizo nation rather than segregated and marginalized. *Indigenista policies* were affected in part through the provision of modern

healthcare and other services. However, many Indian groups wanted to embrace their own identity and culture under a localistic identity. Localistic identity was reinforced by local symbols, such as saints and sacred land. Many indigenous populations wanted to retain their local language and parts of their traditional religions. Despite the desire for localized control of indigenous populations, an assimilation program based on *indigenismo* did succeed in reducing the number of people in Mesoamerica that were still using subsistence farming and speaking indigenous languages (Lauria-Santiago and Binford 2004).

2.2.5 Recent Guatemalan History

In more recent years (1944 AD – present) the history of various Mesoamerican countries offer their own trajectories (Jones 1989). Major political differences in Guatemala have left a lasting impact on the social and demographic structure of the Maya. In 1944, a radical political movement spread throughout Guatemala following a governmental coup against the current dictator, Jorge Ubico. The following decade brought a time of free speech, political movement, the promise of progress, and land reform. This period is referred to as the Ten Years of Spring (1944 AD - 1954 AD), and is viewed by many historians as a period of peace and an acceleration of modernization. However, the proposed land reform policies during this time, which were to redistribute lands from large plantations to landless workers in order to increase productivity, led to a backlash by Guatemalan and U.S. elites (Grünberg and Misión de Verificación de las Naciones Unidas en Guatemala. 2003; Jones 1989). This reform movement was occurring at a time of great

population growth in the countryside, and in order to support their families, many Indians were forced to migrate to seek employment, leading to social unrest among the Indians. Due to a fear of social unrest in Guatemala, and a fear of a president (Jacobo Arbenz) who was sympathetic to communists, the Ten Years of Spring were cut short by a U.S.-backed coup. However, the Ten Years of Spring had already left its mark on the population organization of Guatemala Mayans as land reformation policies affected migration patterns across Guatemala (Jones 1989). Increased population, landlessness, and poverty in the countryside resulted in increased mobility from rural to urban and coastal areas (Grünberg and Misión de Verificación de las Naciones Unidas en Guatemala. 2003; Jones 1989).

Following the Ten Years of Spring, there was continued social unrest, which peaked between 1978 and 1982 with nearly 100,000 individuals massacred. In the midst of the uprisings, a political protest was held by K'iche Mayans at the Spanish embassy in Guatemala City on January 31, 1980. Participants were protesting the mistreatment of Indians by the Guatemalan army, but the protest was cut short by the burning of the Spanish embassy by the Guatemalan police. Many key activists and union leaders were killed in the fire. So, while a Civil War had been ongoing in Guatemala for nearly 20 years, this event is seen as the spark for future rebellion. This instigated a civil war that lasted until 1996, a time period referred to as *la violencia*, in which millions of Indians and Ladinos fled Guatemala to nearby countries or were forced into hiding, and more than 200,000 died (Carmack 1988; Jones 1989). So, while Guatemalan Maya populations were experiencing slow population growth, Mexican Mayan populations were rapidly expanding. For example, the Yucatec population more than doubled between 1950 and 1990 (Sullivan 2000). The Peace Accords were signed in 1996, ending *la violencia* on the international stage, but the destabilization of this country

has resulted in continued violence, mostly directed toward the Indians and peasantry that continues today.

2.3 LINGUISTICS

For the following section, the spelling of language names is often controversial, so for the purposes of this text, spellings suggested by the Ethnologue Index (Grimes 1992) were used except in the case of the Poqomchi'. In this instance, the spelling suggested by Ethnologue, Pocomchí, did not match that used by the population under study. Therefore, I used the spelling reported in the questionnaires filled out by the participants themselves.

2.3.1 Linguistic Classification in Mesoamerica

As mentioned previously, Mesoamerica can be designated as a major region based upon shared culture history and archaeological history. However, there are a number of distinct linguistic traits that also separate the Mesoamerican populations from other American Native groups. While there are more than 100 different languages, each with many different dialects, spoken in Mesoamerican, these languages can be divided into 11 different language families. Language families are delineated based upon shared linguistic features such as phonology, syntax, and word and sentence morphology, and are assumed to represent language groups descended from a common ancestral language (Campbell 1997; Carmack et al. 2007), and

developed in concert during the Preclassic period along with cultural exchange (Campbell 1997). There is some argument among linguists as to the exact grouping of these language families; one proposed grouping at European contact is presented in Figures 2.1 and 2.2. Among these language families, the Oto-Manguen family possesses the greatest variation with more than 40 languages, followed by the Mayan language family with more than 30 (Campbell 1997; Carmack et al. 2007; Coe 2005; Grimes 1992). The Nahua group is actually a member of a larger linguistic family with roots outside of the Mesoamerican cultural region (Campbell 1997; Carmack et al. 2007). There is some evidence that the shared linguistic features of this language group to other Mesoamericans is the result of convergent linguistic evolution or language sharing after a recent intrusion into the area (Carmack et al. 2007). Note that these groupings are highly controversial. In fact, Ruhlen (Ruhlen 1987) claims that Maya is a subfamily of the Penutian language family, which has a wide geographic distribution stretching from Mesoamerica, along the west coast of the United States and into Canada (Ruhlen 1987).

Mesoamerican languages are tied together through many shared features, but these do not necessarily denote a common lineage. Instead, many of these features are thought to be the result of shared contact and, therefore, reflect a convergent evolution of language. Common Mesoamerican phonology (language sounds) is marked by the lengthening of vowel sounds and the addition of glottal stops after many vowels and consonants. Also common among Mesoamerican languages are affixes attached to the verbs that indicate the person and number of persons that the verb is referring to. This word morphology also takes into account the social status of the person(s) of interest. Syntax, or sentence structure, is variable across many of the populations, but a common subunit of the Mesoamerican sentence construction is

the verb-before-object order. Also, the common units of count are in groups of 20 rather than counts of 10 common in modern Western culture. The best evidence for convergence of these language families is the presence of shared or loaned idiomatic expressions. These expressions often relate to the shared calendar system (e.g. terms for star, Venus, sun) (Campbell 1997; Carmack et al. 2007). These traits are almost universally shared among the languages of the Mesoamerican region; however, this classification has caused some languages whose speakers share similar cultural traits to be excluded from the Mesoamerican language area, including the Lenca, Xinca, and Chorotega.

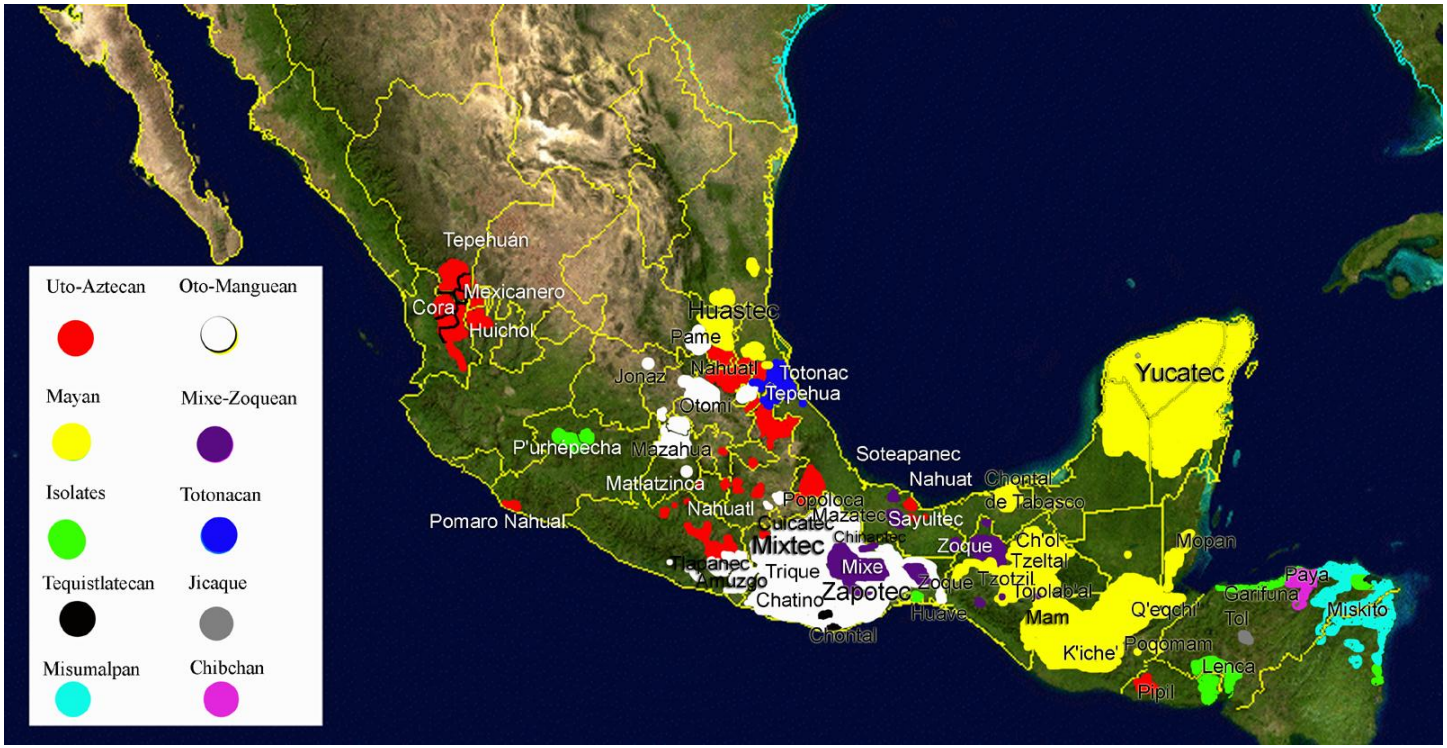

Figure 2. 1. Map of the distribution of languages within the Mesoamerican cultural region (image part of the public domain, available through wikicommons at http://en.wikipedia.org/wiki/File:Mesoamericanlanguages.png).

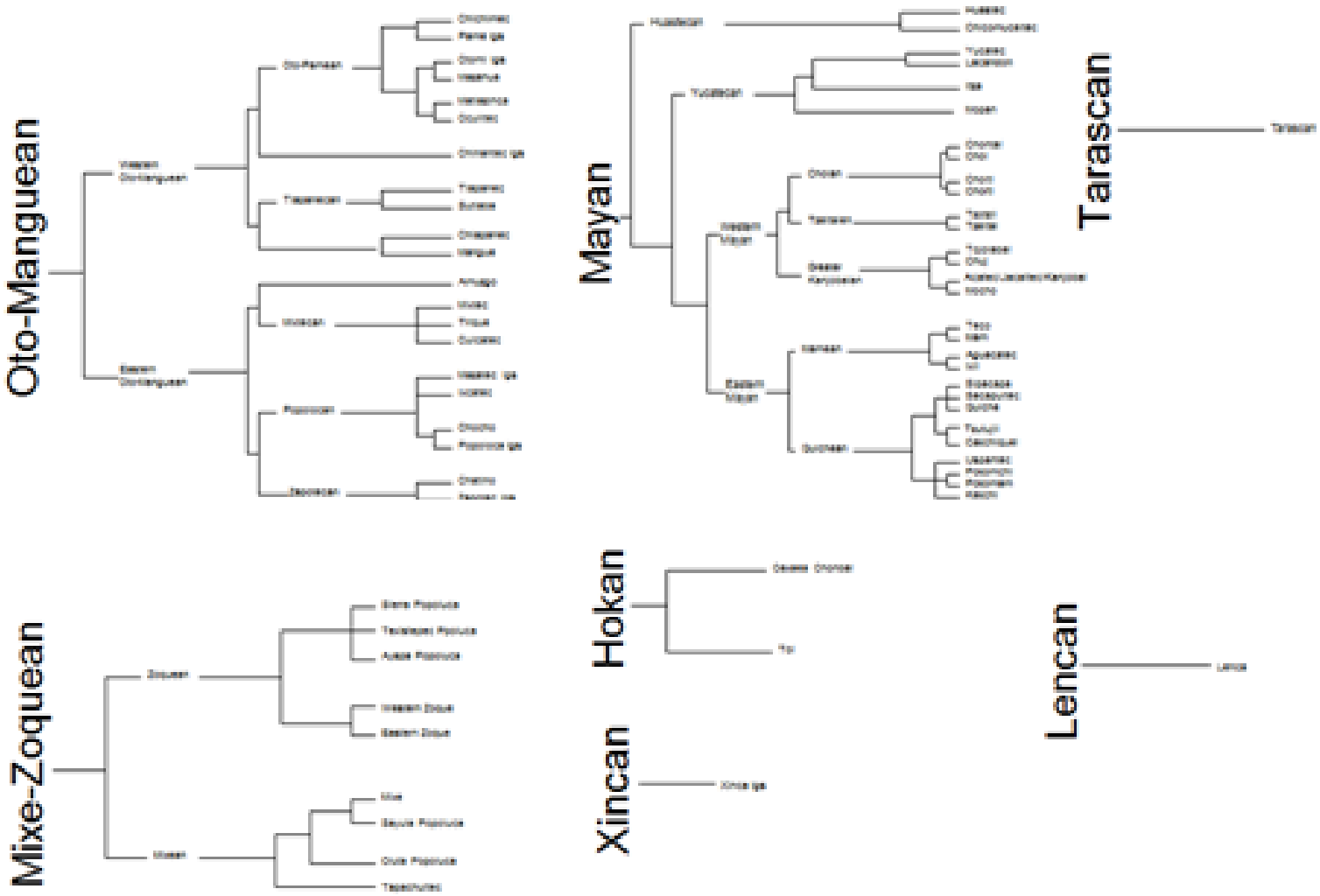

Figure 2. 2. Relationship of Mesoamerican languages within the Mesoamerican cultural region (adapted from Carmack 2007, Campbell 1997).

2.3.2 Mayan Language Family

The Maya family contains approximately 30 distinct languages which can be found in Mexico, Guatemala, Honduras, and Belize (Coe 1966; Ruhlen 1987). Linguistic evidence indicates that the first Maya speakers moved south across Mexico and settled in the western Guatemalan highlands more than 5,000 years before present (yBP) (Campbell 1997; Coe 1966). Using glottochronology, a dating technique that assumes that language grammar diverges at a constant rate, the first language group to split was the Huastec (Wastek) 4000 BP, later moving north to Mexico to be surrounded by Nahua speakers along the eastern Mexican coast (Campbell 1997; Carmack et al. 2007; Coe 2005; Vogt 1969). Figure 2.3 displays the phylogenetic grouping of the Maya languages using these techniques. Over the next few thousand years a few major language branches moved down from the highlands to occupy the Petén (central Maya regions) and Yucatan (Northern Maya Area) and further diversify. These Yucatec and Ch'olan branches diverged next, spreading to the Northern and central Maya regions around 3000 BP. Next, core Quichean and Mam diverged in the highlands around 2000 BP. Then western Guatemalan languages in the southern highlands remained more conservative over time so many of the Mayan languages spoken today in Guatemala diverged only 1-2,000 BP (i.e., K'iché, Poqomchi', Kaqchikel) (Campbell 1997; Coe 2005; Vogt 1969). While written language and spoken language are often considered separate entities by linguists, glotto-chronological analyses of the relationships among Mayan languages fit very well with the recorded history from texts, much better than among other language families in the Americas (Carmack et al. 2007). The divergence of these language groups explains much of

the current geographic distribution of these languages and correlates well with archaeological evidence on these populations, as illustrated in Figures 2.4 and 2.5 (Campbell 1997; Kaufman 1976).

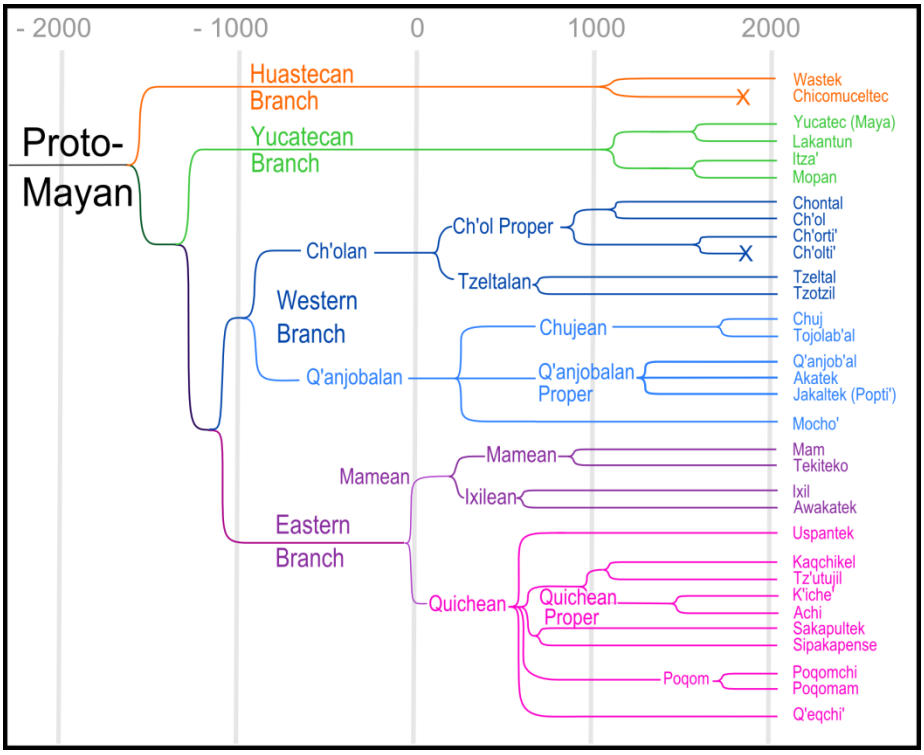

Figure 2. 3. Relationship among Maya languages based on analyses by (Campbell 1997; Watanabe 2001). Figure made available through Wikicommons (http://en.wikipedia.org/wiki/File:Mayan_Language_Tree_in_colour.png)

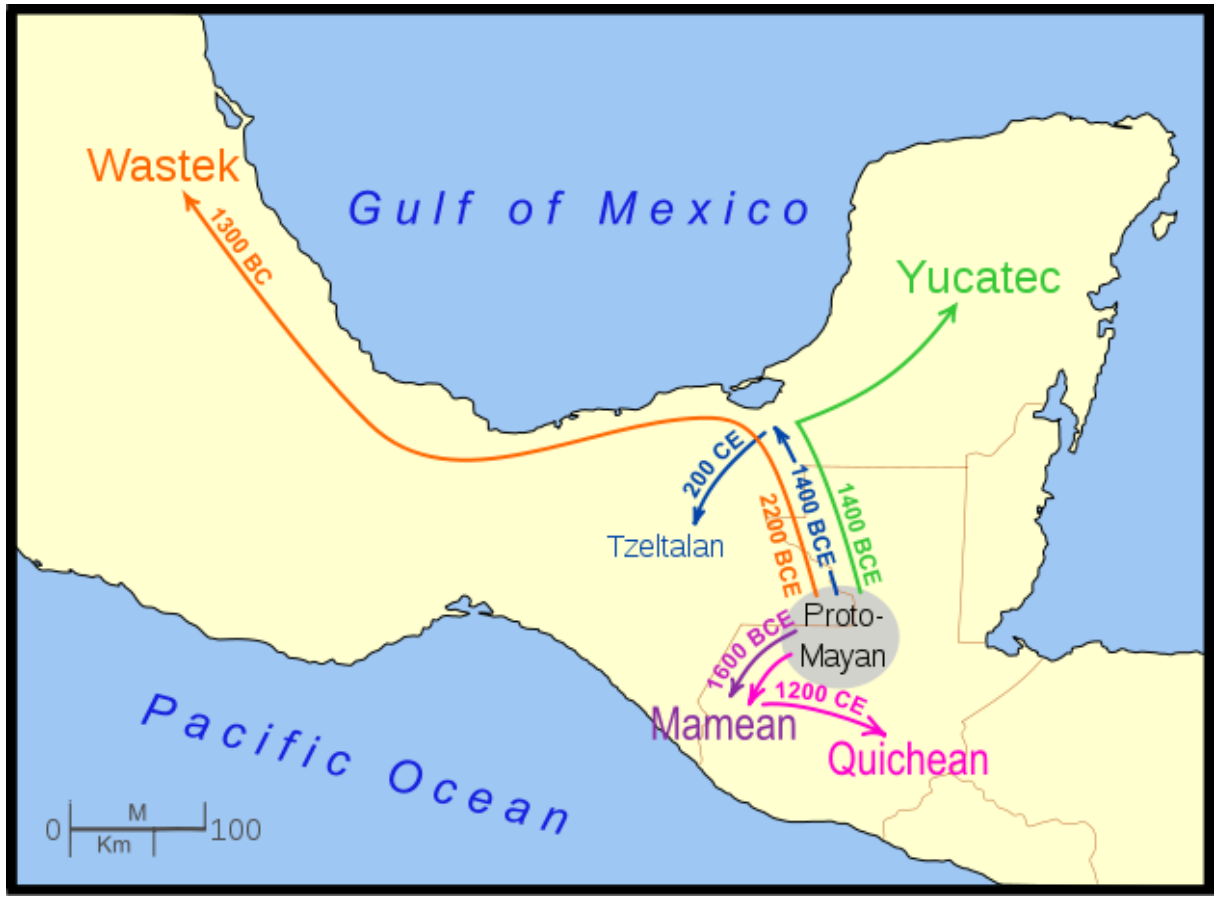

Figure 2. 4. Map of s thedispersal of Mayan populations based on glottochronology. Figure made available through Wikicommons (http://en.wikipedia.org/wiki/File:Mayan_Language_Migration_Map.svg).

Figure 2. 5 Current geographical locations of the Maya populations based upon language. Figure made available through Wikicommons (http://en.wikipedia.org/wiki/File:Mayan_Language_Map.png.

The same process of convergent evolution that operated on Mesoamerican languages as a whole also operated on the Mayan language family. In the distant past this was evident in the hieroglyphic writing system. The Mayan lowland languages possess the greatest number of hieroglyphs in Mesoamerica, but are written primarily in one language, Ch'olan. However, in the Yucatan Peninsula, texts exist written in a mixed form of Ch'olan standard with Yucatecan

language rules and some Yucatec vocabulary (Campbell 1997; Carmack et al. 2007). Today language sharing among the Maya is most evident among the Q'eqchi'. This language has a wide distribution across the northern regions of Guatemala, but the dialect spoken in Cobán is considered a "prestige dialect." Therefore, new linguistic trends usually begin in this city and spread through Q'eqchi' speakers throughout Guatemala (Carmack et al. 2007). This pattern of convergent evolution among Mayan languages is one reason why they have been treated as a homogenous group. While Mayan languages are still widely spoken today across Mexico, Guatemala, Belize, Honduras, and El Salvador, due to Spanish or English as the governmental language, few Mayans remaining are monolingual and this usually only includes women (Watanabe 2000).

2.4 BIOLOGY OF THE MAYA

2.4.1 Morphology and Classical Genetics

Many expeditions were made to collect physical data on Native American populations during the early 20th century. Early investigations at the turn of the century focused on anthropometric measurements, but from the 1940s to the 1960s investigations expanded to include blood group markers, nutritional, and demographic data. Some generalities about the populations of Mesoamerica can be made using these early data (Faulhaber 1970; Justice et al. 2010; Matson 1970; Roberts and Newman 1970).

In many of the early investigations, measurements were taken while participants were still wearing clothes. Some investigators corrected for mis-measurement when looking at weight by subtracting the weight of the standard dress in that population. However, this was not done routinely, and often corrections were not made for other measurements, such as stature (Faulhaber 1970). Therefore, any interpretation of these data should be done carefully, but a few generalities can be made about findings across many of the studies despite these difficulties. First, within groups, a small secular trend toward increased stature and increased height of the face has been noticed between measurements taken at the turn of the century compared to those taken middle 20th century. These differences were less marked among the so-called "pure" Indians, and have therefore been considered a result of European admixture (Faulhaber 1970). Among groups a cline in stature has been noted that runs from greater stature in northern Mexican Indians and mestizos, shorter stature in the southern region of Mesoamerica, including the Chiapas, Guatemala, and Honduras; with some evidence of a slight increase in stature south of the Honduras/Guatemala border. The shortest stature among all populations measured during this time period is found among the Guatemalan and south Mexican Maya. There are isolated pockets of short stature found in northern Mexico, but these do not greatly alter the strong correlation between latitude and body morphology (Faulhaber 1970; Justice 2007; Justice et al. 2010). Much of the change in height is due to a reduction in the length of the extremities, as Mayans have roughly equivalent sitting height to other Mesoamericans and Central Americans, but shorter overall stature.

Differences in head and face shape have also been noted among populations of Mesoamerica and Central America. First, the cephalic index, which compares head length and

head breadth, indicates that populations in northern Mexico and along the Gulf coast tend to be mesocephalic (intermediate cephalic index or slightly long). In contrast, the populations along the lower periphery of Mesoamerica including the northern portion of Central America and the entire Maya region have a tendency toward brachycephaly (higher cephalic index or more rounded skulls). These same regions also tend to have proportionately longer faces with higher and narrower noses than found in north Mexico (Faulhaber 1970) This continued pattern of variation in anthropometrics across latitude has also been shown when combining anthropometrics in multivariate analyses (Justice 2007).

A recent study by Scherer (2007) used the dental remains of 827 Mayans from 12 different Classic period Maya sites to look at the effects of genetic drift and gene flow on odontometrics. This study found little population substructure among the lowland Maya populations from the central and northern Mayan regions, but there was a slight separation of the one southern site of Kaminaljuyu (currently Guatemala City). However, the overall population structure among these sites was low (F_{ST} = 0.018). Furthermore, there was little evidence of non-Mayan gene flow or genetic drift as populations tended toward expected values of variation (Scherer 2007).

Blood group marker data were collected throughout the 20th Century on a number of populations from Meso- and Central America including both “pure” Maya and “mestizo” Maya from Mexico and Guatemala. After combining data from several studies, Matson (1970) has summarized a few generalities among the blood group markers. First, all populations identified as “pure” Indian were near fixation for the frequency of the O allele in the ABO system. This was apparent in the Maya (including Yucatec, Huastec, K’iche, Tzotzil, Chol, Itza, Lakandon,

Tzeltal, Poqomam, and Mam) populations, where the lowest frequency of the O allele across all Mayan was 0.7677, identified as a Spanish-speaking Mayan population in Mexico (Matson 1970). This cultural area also shares a high frequency of *M* and low *N* for the MNS system. The Rh system provides further evidence of little population substructure among Mesoamerican populations, with consistently high frequencies of the *CDe* and *cDE* haplotypes, but almost absent *cDe* (high in African populations) and *cde* (high in European) haplotypes (Matson 1970). A few other blood group markers, although sampled less extensively, offer estimates of admixture into Mesoamerica from Europe and Africa. For example, the presence of the *V* antigen for the Rh system is indicative of African admixture and is primarily found in populations close to the Caribbean coast, and the presence of *A*, *B*, for the ABO system and *Kell* antigens has been viewed as evidence for European admixture (Matson 1970). However, the assumption that *A* and *B* alleles are the result of European admixture may be unfounded, as these alleles have been found in high enough frequencies among native populations in North America to make admixture an unlikely explanation in all cases (Crawford 1998).

Finally, physiological studies in Mesoamerica have revealed a difference in metabolism between the Maya and the rest of Mesoamerica. The Maya exhibit higher metabolic rate, lower pulse rate, lower blood pressure, and higher incidence of hyperthyroidism (Roberts and Newman 1970). Also, while their reliance on a primarily processed corn and bean diet limits damage caused to their teeth, this diet subjects the Mesoamericans, and especially the Mayans, to several nutrient deficiencies including Vitamins A, C, E, iodine, and protein (Roberts and Newman 1970; Scherer 2007).

2.4.2 Molecular Genetics

2.4.2.1 mtDNA

Mitochondrial DNA (mtDNA) is a useful molecular tool for answering anthropological questions about human evolution and variation due to the inherent characteristics of mtDNA inheritance (Avise et al. 1987). mtDNA is almost entirely inherited through the mother, thus providing a deep maternal history. Mitochondrial DNA does not undergo any recombination and has a constant mutation rate. Therefore, any changes in genetic markers seen across generations are assumed to be the result of new mutations, making this system ideal for statistical analyses. The rate of mutation can be used to develop a chronometer that estimates the separation of founding and offspring populations. Additionally, mtDNA resides within the cell's mitochondria, which are readily abundant, making mtDNA easier to obtain from the cell than nuclear DNA. Given its utility for anthropological analysis, mtDNA is often used to reconstruct population history in the Americas and for contrasting the patterns of genetic diversity within and among American Indian populations. Furthermore, mtDNA presents distinct haplogroups (inherited groups of genetic markers which represent a discrete lineage) that differ among geographically distant populations. These distinct haplogroups in turn allows for comparison among worldwide populations. The difference in mtDNA lineages between continents allows researchers to measure the amount of maternal gene flow through migration

(e.g., proportion of Native American, European, Asian, and, African mtDNA haplogroups found in the Americas).

Five major mtDNA haplogroups characterize Native American mtDNA diversity: A, B, C, D, and X. All five of these haplogroups are present in Siberian and Asian populations, and vary in frequency across the Americas (Figure 2.6). Haplogroup A, defined by a mutation at nucleotide position (np) 16111, involving a C to T transition and including a *Hae* III enzyme restriction at np 633, has high frequencies in Alaska, Canada, eastern portions of the US, and Mesoamerican populations. Haplogroup B is defined by the presence of the 9-bp Region V deletion. Haplogroup B is found in high frequencies among the indigenous peoples of the western and midwestern US and is almost absent in Arctic populations. Lineage C is defined by the loss of a *Hinc* II restriction site at np 13259 and the gain of an *Alu I* restriction at np 13262. Haplogroup C is rare in most Native North Americans but increases in frequency across South America. D is the loss of an *Alu* I restriction site at np 5176. Haplogroup D has been shown to occur in higher frequencies among Native Alaskans, in lower frequencies in the rest of North America, and in high frequency among South American populations residing in Amazonia (Bonatto et al. 1996; Bonatto and Salzano 1997a; Bonatto and Salzano 1997b; Lalueza-Fox et al. 2001; Lalueza-Fox et al. 2003; Mateus Pereira et al. 2005; Merriweather and Kaestle 1999; Merriwether et al. 1995; Rubicz et al. 2003; Salzano 2002; Sandoval et al. 2009; Schurr 2004; Schurr et al. 1990; Schurr and Sherry 2004; Smith et al. 1999; Torroni et al. 1993a; Torroni et al. 1993b). Involving a T to C mutation at np 16189, a C to T at np 16278 and an addition of *AccI* at np 14465, haplogroup X is found in high frequency around the Great Lakes and in Greenland, with moderately lower frequencies elsewhere (Rubicz et al. 2003; Schurr 2004; Schurr and Sherry 2004).

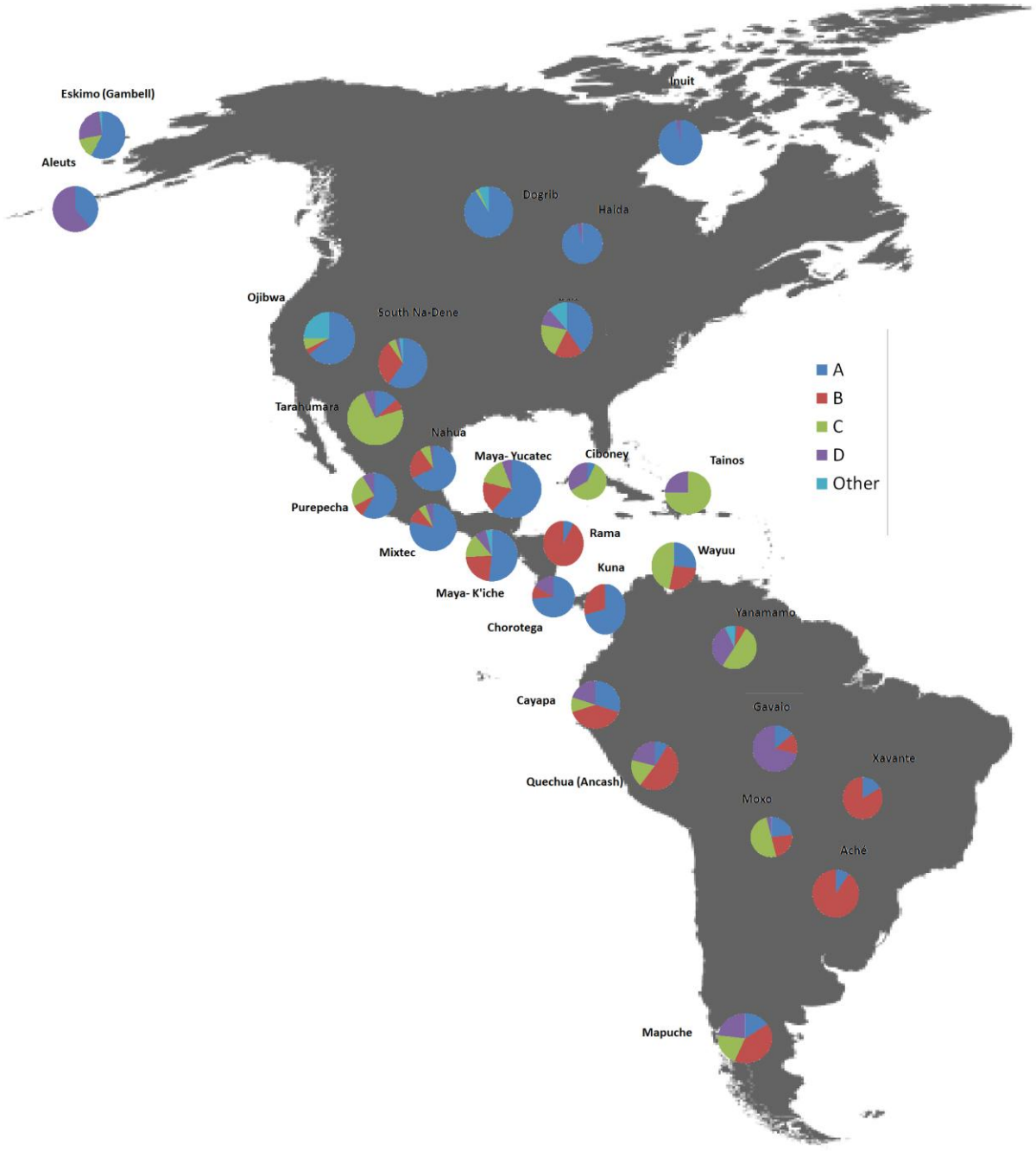

Figure 2. 6. Frequency of major mtDNA haplogroup for the Americas (created by author with data from comparative populations in this study, Rubicz 2007, and Torroni et al. 1993).

Since the seminal work of Vigilant et al. (1991), study of human mtDNA sequence variability has become commonplace and has allowed for greater resolution of mtDNA haplogroup relationships. Mutations specific to circumpolar populations and other Native Americans for haplogroups A2, C1, D1, and X2a have been found in the hypervariable regions (HVS1 and HVS2) of mtDNA, which, when combined with RFLP analysis, reveal possible founding lineages for the New World. A2 is characterized by the following HVS1 sequence motif 16111T, 16223T, 16290T, 16319A, and 16362C as compared to the Cambridge Reference Sequence (CRS). Haplogroup X2a is characterized by the addition of a mutation at np 16183C and np 16213A. Haplogroup B2 combines the 9bp deletion with a 16182C, 16183C, 16189C, 16217C sequence motif. C1 (16223T, 16298C, *16325C*, 16327T) and D1 (16223T, *16325C*, 16362C) are both differentiated from the parent haplogroups through the mutation at np 16325 (Achilli et al. 2008; Sandoval et al. 2009).

Using patterns of variability in mtDNA haplogroup and sequence variation, researchers have made inferences into the peopling of the Americas and the historical relationships between Native American populations. Native Americans overall exhibit decreased variation compared to Europe, Africa, and Asia, and variation decreases moving from North America to South America (Tarazona-Santos et al. 2001; Wang et al. 2007). Controversy surrounds the number and timing of migrations into the New World, but it is generally accepted that the Americas were peopled from Siberia sometime before the last glacial maximum ~20,000 yBP (Achilli et al. 2008; Wang et al. 2007), with an entrance into South America by ~13,000 yBP (Fuselli et al. 2003).

As mentioned above, haplogroups A, B, C, D, and X2a comprise the whole of Native American mtDNA diversity, and are only shared with Siberian and Asian populations, the presence of any other haplogroups are an indication of gene flow. European admixture can be detected through the presence of haplogroups H, I, J, K, T, U, V, W, and X, while African haplogroups are L0, L1, L2, and L3 (Salas et al. 2004; Schurr 2004). Studies such as the one described here provide evidence for the impact of European colonization and the slave trade on the genetic makeup, as well as the cultural identity, of populations in Latin America.

2.4.2.3 Y Chromosome

Like mtDNA, the Y chromosome markers are uniparental, but are passed on from father to son rather than from mother to daughter. The Y chromosome is approximately 58 million base pairs in length, and includes large segments that do not recombine with the X chromosome. There are small pseudo-autosomal fragments close to the ends of the chromosome that are subject to recombination, but represent only 5% of the total chromosome (Hammer and Zegura 2002; Jobling et al. 2004). Due to the absence of recombination, much of the Y chromosome is highly conserved and therefore provides information on deep paternal lineages. Many SNPs, insertions/deletions (indels), have been detected that have been passed down through many generations and therefore can be used to trace lineages or haplogroups. Like mtDNA haplogroups, Y haplogroups cluster in geographical regions and can be characterized by RFLP, sequencing, or probe assays. Y chromosome SNP

variation is greater among populations and among continental regions than mtDNA or autosomal STR markers, with Y SNPs exhibiting an average of 52.7% variation among continents (Seielstad et al. 1998). This attribute makes Y SNPs uniquely informative on male migration patterns.

Due to the inconsistency with which these markers are used and named the Y Chromosome Consortium (YCC) was established to standardize the characterization of these mutations, first published in 2002 (Y Chromosome Consortium 2002), and updated again in 2008 (Karafet et al. 2008). The Y haplogroups are named by letters A through T along with SNP designation, with subhaplogroups within these major lineages named by combinations of letters and numbers and SNP name. As a result, use of the Y chromosome SNPs and indels for looking at population structure has increased greatly within the past two decades.

In addition to SNPS, the Y chromosome carries several repetitive sequences of DNA that, like sequence data in mtDNA HVS1, provide higher resolution data on population history and mutate more rapidly than SNPs. Those most commonly used are short tandem repeat markers (STRs) containing 3-8 base pair (bp) repeats. While there are nearly 500 STRs on the human Y chromosome (Kayser et al. 2004), there are 16 markers more commonly used due to their availability in commercial typing kits, including H4, DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, and DYS635 (Applied Biosystems, Inc.). The cost of these kits is sometimes prohibitive, and the kits have only recently become available, therefore, these markers are still not consistently used, making comparisons across populations challenging.

Investigations of populations in the Americas have found only two haplogroups, Q-M242 and C-M130 that contributed to the original founding populations, but which originated in Central Asia. Of these, haplogroup Q-M242 and its subgroups are the most common and can be found in all population tested within the Americas (Figure 2.7). Haplogroup C-M130 is common among populations in northwestern North America and almost absent in Central and South America (Malhi et al. 2008). Haplogroup C-M130 has a total of 19 lineages identified by 30 mutations, but not all of these can be found in the Americas. Haplogroup Q-M242 has 13 sub-haplogroups marked by 17 SNPs according to the last YCC update (Karafet et al. 2008); however, the International Society of Genetic Genealogy (ISOGG) just updated their topology to include recently discovered SNPS displaying 22 sub-branches of haplogroup Q with an additional 18 SNPs not included on the 2010 haplogroup tree (www.isogg.org, 2011). Haplogroup P-M45 (also known as R-M45) is also common in New World populations with frequencies ranging from 4% - 88%. The highest frequencies of R1 are found in northeastern North America, but this haplogroup is absent in eastern Asia, so is now considered the result of recent admixture with Europeans (Malhi et al. 2008).

Additional, SNPs are used to identify mutations shared exclusively among populations in the Americas. These include Q1a3a1-M3 (also known as Q3 and more commonly Q1a3a-M3) and C3b-P39 (Karafet et al. 2008; Zegura et al. 2004). Native American Q can further be subdivided into three subclades, Q1a3a1a-M19, Q1a3a1b-M194, and Q1a3a1c-M199, with the last subclade also having mutations at P106 and P292. Haplogroup C3b currently has no known subclades unique to the Americas (Geppert et al. 2011).

Several attempts have been made to estimate the time of expansion from the most recent common male ancestor carrying haplogroup Q-M242 and Q1a3a1-M3. Dating methods include estimates derived from SNP only models, STR models, and also comparing the mutation rates and diversity among both SNPs and STRs within these clades. Expansion dates using Q-M242 range from ~18,000 to ~15,000 BP (Bortolini et al. 2002; Bortolini et al. 2003; Schurr 2004; Zegura et al. 2004). Using Q1a3a1-M3, which likely represents the demographic or spatial expansion of males after entering the Americas, dates range from ~30,000 to ~7,600 BP (Bianchi et al. 1998; Forster et al. 2000; Hammer et al. 1998; Karafet et al. 1999; Underhill et al. 1996). This range decreases and converges on ~13,800 BP when only SNPs are considered. These dates still support the claim that Native Americans entered the continent sometime before the last glacial maximum (Schurr 2004). Haplogroup C-M130 is likely much older than haplogroup Q-M242, with dates ranging 30,000 to 25,000 BP using SNP only models (Karafet et al. 1999; Schurr 2004; Underhill et al. 2000).

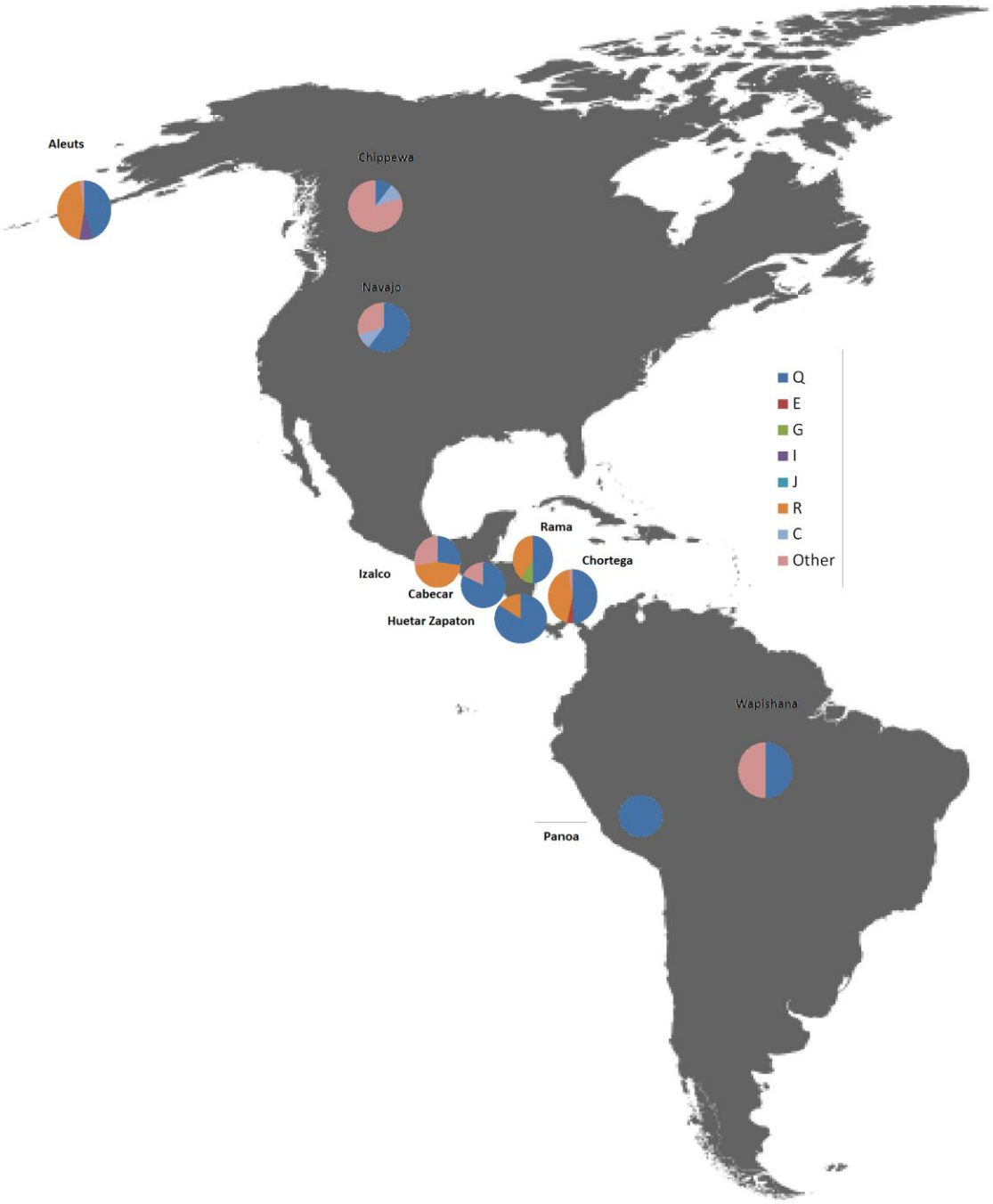

Figure 2. 7. Frequency of major Y chromosome haplogroups for the Americas (drawn by author using comparative samples from this study, Rubicz 2007, and Schurr 2004).

2.5 POQOMCHI' MAYA

Poqomchí is a Mayan language distributed across two Guatemalan departments (similar to U.S. counties), the Alta Verapaz and Baja Verapaz, and seven municipalities. The number of speakers is rapidly declining with an estimated 90,000 Poqomchi'-speaking Maya in 1987, and only 50,000 speakers in 1997 (Ruhlen 1987, Cahuec del Valle 1997). This language is most closely related to Poqomam and the Poqom languages that are closely related to the Core Quichean languages and have a slightly more distant relationship to the Q'eqchi' Maya, with which they overlap in geographic distribution (Ruhlen 1987). Poqomchi' first diverged from Poqomam due to an intrusion of Q'eqchi'-speakers around 700 BP, and were subsequently pushed into former Xinca territories in eastern Guatemala around the Polochic River (Cahuec del Valle 1997; Campbell 1997).

This area was part of the Maya civilization before the Spanish arrived in the 1520s and subjugated the highlands of Guatemala. The Poqomam, Q'eqchi', and Poqomchi' had only recent gained their freedom from K'iche' rule when the Spanish first arrived in 1524 (Olson 1991; Reina 1969). The Maya of Alta Verapaz were able to hold off the Spanish conquistadors. However, the natives were soon peacefully converted to Christianity, giving Spanish control over the people (Carmack 1986). The area, once referred to as Tuzulutlán, or Land of War, was named "Verapaz", meaning "True Peace" in Spanish, after this peaceful conversion. The Spanish further subdivided the Poqomchi' pushing half of the population east, including populations currently in Tamahú, Tactic, Purulhá, and Tucurú, and the other half to the western

municipalities of Santa Cruz, San Cristóbal and Belejú. During this move the populations were evangelized. It is unknown how many members of surrounding Maya were included in this reduction, but the Poqomchi' resided among K'iche, Kaqchikel, Chol and Poqomam during the Postclassic period (Cahuec del Valle 1997). Since the Dominicans peacefully converted the Poqomchi' and surrounding highland population of the Verapaz, many more men survived in this area and families were able to stay on their land (Carmack 1986). Due to their greater survival, slaves were not needed as a labor force in this region and the Indians continued to outnumber Europeans and Ladinos. After colonization, population size quickly recovered in the highlands (Lovell and Lutz 1995; Reina 1969; Watanabe 2000).

In the 19th century the area became an important coffee-producing region, which led to land disputes between German settlers and investors, the Ladinos, and the more traditional natives occupying the land. Coffee is still an important crop in this region, but very little land remains in the hands of the Poqomchi'. Landlessness and the growing population of Q'eqchi' Maya in the Polochic River Valley has caused disagreements between municipalities and among aldeas within the municipalities (Cahuec del Valle 1997). The Guatemalan government responded to the Indian demands with force, and during the Civil War of the 1970s and 80s, there was a great reduction in population size in Alta Vera Paz, including among the Poqomchi', due to violence and migration to Mexico and the U.S. (Carmack 1988; Lovell and Lutz 1995; Olson 1991). Like after colonization, the population size of the Alta Vera Paz recovered rapidly following these hardships (Figure 2.8). Nevertheless, their history has led to their marginalization and relative isolation, and as a result, there is a relative lack of research available on the Poqomchi' as compared to other surrounding Mayan populations.

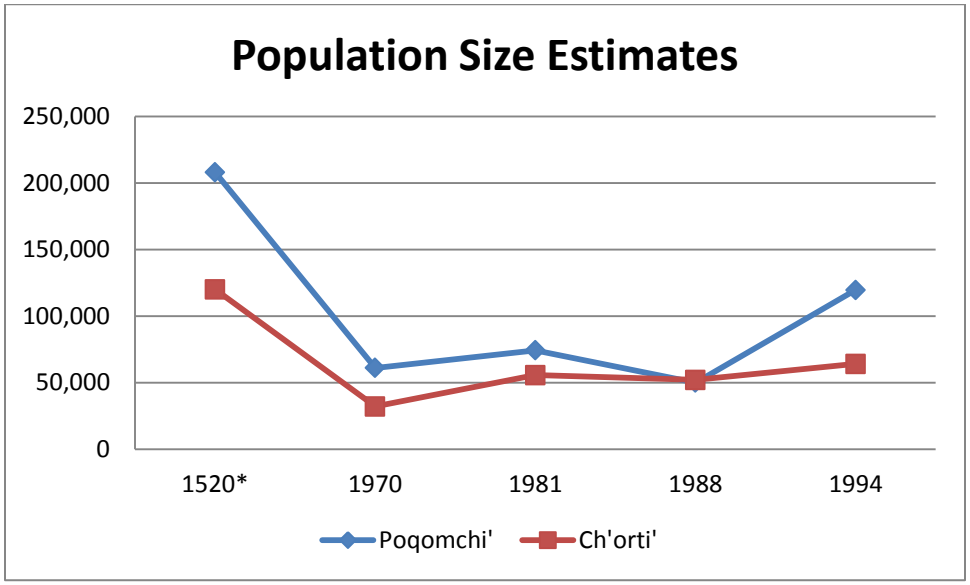

Figure 2. 8. This plot provides population size estimates by year for the Poqomchi' and Ch'orti' Maya regions. For 1520, there is no population size estimate available for just the Poqomchi', so the estimate for the entire Verapaz is provided, which would include the Qeqchi' Maya (Lovell and Lutz 1995).

2.6 CH'ORTI' MAYA

Ch'orti' Maya is spoken mainly in eastern Guatemala, but can also be found in western Honduras and was once spoken in northwestern El Salvador. The Ch'orti' language descends from the Cholan branch of Classic Maya, which split into the Ch'olti' and Ch'orti' in eastern Guatemala. Proto-Ch'orti' may have first been spoken in the central Guatemala highlands around Kaminaljuyú, but climatic changes and other advancing Maya groups caused populations to migrate to the east (Sharer 2009), where descendants of Ch'orti' speakers

occupied this region for almost 2,000 years (Metz 2009). At least two major Classic Period centers existed in Copán, Honduras and Quiriguá, Guatemala (Evans and Webster 2001; Metz 2009; Sharer 2009). These centers were important for trade as they acted as a communication point between the northern and southern Maya areas, as well as provided much desired jade and obsidian (Evans and Webster 2001). While it is clear that the Maya were the ruling class, there is also evidence that the Lenca, Xinca, or other non-Maya groups may have been residing around the Maya centers in the east. Analysis of Classic Mayan glyphs indicates that the primary language of the ruling class in the Central Maya area was proto-Cholan, which later gave rise to Ch'orti' (Evans and Webster 2001; Manahan 2008; Sharer 2009).

When the Spanish invaded the Ch'orti' area in 1524, there were approximately 120,000 Mayans inhabiting the region (Lovell and Lutz 1995; Lutz et al. 1982; Metz 2009). While the populations united against the Spanish, they were eventually defeated and pushed off their most fertile lands or enslaved. The Spanish invasion was a series of attacks carried out with the help of subjugated indigenous Mexicans and Mayans that lasted from 1524 AD until 1535 AD (Torres Moss 1996). The warfare, out-migration, and disease brought by the Spanish, precipitated a population decrease of as much as 90% in some parts of the Ch'orti' region (Metz 2006; Metz et al. 2009). Those Ch'orti' who survived were relocated in Guatemala to today's Jocotán, Camotán, Olopa, San Juan Ermita, La Unión, Quetzaltepeque, and Esquipulas (Grünberg and Misión de Verificación de las Naciones Unidas en Guatemala. 2003; Olson 1991; Reina 1969), as well as several towns in Honduras and El Salvador. Europeans and Ladinos were more likely to move to cities with important commodities and along trade routes. The Ch'orti' region was bounded by two major trade routes and was thus of some importance to the

Spanish. Indigo, sugar, other cash crops, and mining also attracted settlers to the region (Metz et al. 2009). The best well-watered valley lands in this area were given to the Ladinos and Europeans instead of the indigenous populations (Grünberg and Misión de Verificación de las Naciones Unidas en Guatemala. 2003). Due to the rapidly declining Indian population, the Spanish began to import African labor in Meso- and Central America already in the first decades of the conquest period. Starting in the second half of the 19th century, a spiral of population increase, by both Ch'orti's and Ladinos, along with land privatizations and greater inequality slowly undermined Ch'orti' subsistence, leading them to demand more land and political participation. Aggravating the volatile situation was extreme racism in the region and the Ladino ability to co-opt state power to dispossess Ch'orti's of their land and labor. Ch'orti' attempts to organize were met with state terror, causing several waves of exodus to Honduras and even Mexico and the U.S. form the 1950s to 1980s, with as many as one out of every four Ch'orti Maya fleeing their homes for safety in Mexico and the U.S. (Metz, Personal communication; Olson 1991). As a result of Spanish settlement, the introduction of African slaves in the 1500s, the *Ladinoization* of many Indians taking refuge in haciendas and towns to avoid race-based exploitation, and sexual predation of Ladino men on Ch'orti' women, there was cultural and biological admixture among the Indians, Africans, and Europeans.

The Ch'orti' region is of particular interest to biological anthropologists as it lies on the fringe of the Maya territory adjacent to the Central American region dominated by non-Mayan lowland populations such as the Chocoan and Chibchan-speaking groups (Cooke 2005; Melton 2008; Metz et al. 2009). Also, it is likely that the history of this region allowed for a higher degree of admixture than found in other Maya regions. Finally, while there exists linguistic,

ethnographic, and archaeological research in the Ch'orti' region, there is a lack of biological data on the Ch'orti' or any descendants of the Classic Cholan Maya. Also important, the Ch'orti' Maya in eastern Guatemala represent the most likely descendants of the Central Maya region remaining in Guatemala, and there are no molecular data available for any ancient or living descendants of the central Maya area (Coe 2005; Metz 2009).

In summary, the Maya are a culturally and linguistically heterogeneous group of populations, which are often treated as a homogenous group biologically distinct from other Latin American populations. The Maya groups began diversification ~4,000 BP, have a long history of permanent and trans- migration. The Ch'orti region lies on the fringe of the Maya territory adjacent to the Central American region dominated by Chibchan language groups. Also, differential amounts of gene flow existed between Ch'orti' and Poqomchi' populations and other non-Maya natives, Europeans, and Africans. The Ch'orti' were conquered, while the Poqomchi' were peacefully converted. There is a lack of biological data on both of these populations and any descendants of the central Maya region. Therefore, both populations offer a unique opportunity to examine the effects of migration and colonization on the Maya, and address important regions missing from the body of biological data used to characterize the variation of the Maya.

CHAPTER 3. MATERIALS AND METHODS

This chapter describes the sample collection methods for fieldwork conducted in the Departments of Alta Verapaz, and Chiquimula, Guatemala in the summer 2007 and winter 2009/2010, respectively. The following laboratory methods are described for DNA extraction from buccal swabs, mouthwashes, and blood samples; mtDNA sequencing and haplogroup assignment; and Y chromosome SNP and STR analysis. Finally, analytical methods are explained including nucleotide diversity and gene diversity to examine within group variation; Analysis of Molecular Variance (AMOVA), Median-joining network analysis, Kimura-2p distances, Multi-dimensional Scaling (MDS), and Neighbor Joining Tree (NJT) to examine among group variation; R-matrix analysis regressed on diversity, mismatch analysis, Tajima's D, Fu's F_s to explore the effects of natural selection, gene flow, and population fluctuations on genetic structure; and finally, Monmonier's Maximum Difference algorithm, SAMOVA, and Mantel randomization tests were performed to look at the relationship between genetics, geography, and language.

3.1 MATERIALS

New data collected for this project focused on two populations, the Poqomchi'-speaking Maya and the Ch'orti'-speaking Maya. Samples were collected during the summer of 2007 from the Poqomchi' Maya of Tamahú in the Alta Verapaz, a department in the north central region of Guatemala (Figure 3.1). Tamahú is a municipality within the Alta Verapaz located in the

Polochic River valley and borders the municipalities of Tactic in the west and Tucurú in the east. In Tamahú there are 26 hamlets; of which 10 are Q'eqchi' and 16 are Poqomchi'. In the winter 2009/2010, samples were collected in Jocotán, Department of Chiquimula, Guatemala from Ch'orti'-speaking Maya, and Ch'orti' descendants. Oral consent was obtained from each participant (HSCL # 16735 and 15165) (See Appendix A and B for the oral consent forms used), and then contact information and project description was provided for each participant in Spanish. Data collection in Tamahú was conducted in concert with an on-going ethnographic study on the etiology of anemia, led by doctoral student James Herynk of the University of Kansas and the local Ministry of Health. Participants were recruited by word of mouth, and the study was announced by volunteers from the Ministry of Health. Participants for the genetics portion of the project were invited to two collection sites during five days, two days in the highland hamlet of Onquilha' and three days at a Ministry of Health satellite clinic located along the Polochic River in the hamlet of Chimilon.

Data collection for the Ch'orti'-speaking Maya took place in Jocotán, the capital of the Jocotán municipality in the Department of Chiquimula, Guatemala. Again, permission and cooperation were received from the Ministry of Health. Participants were recruited by word of mouth and invited to participate at the local Ministry of Health clinic and at the local church community center. Data collection took place during three days. The project goals were first explained to each volunteer, and oral consent was obtained from each participant (HSCL # 16735 and 15165) (See Appendix B for the oral consent form used), and then project description and contact information was provided for each participant in Spanish.

Figure 3. 1. A map of the collection sites for the current study and the approximate sites for the comparative samples used in mtDNA analysis from the Caribbean, Mesoamericna and Central American.

Figure 3. 2. A map of the collection sites for the current study and locations of the comparative samples used in mtDNA analysis located in South America and lower Central America.

Each participant was asked to provide two buccal cell samples (cheek swabs) and one mouthwash sample (saliva). Additionally, Poqomchi' participants were asked for three ml of whole blood. All participants filled out a questionnaire designed to collect information regarding number, age and survivorship of children; number, age and survivorship of relatives at least three generations in depth (i.e., including siblings, offspring, parents, and grand-parents); and migratory information (current residence, birthplace, birthplace of parents and grand-parents). The fieldwork conducted in 2007 was supported by funds from the KU Latin American Studies Fieldwork Grant (2007) and a KU Department of Anthropology Carroll D. Clark Research Award (2007). Fieldwork conducted in summer 2009-winter 2010 was funded by the Tinker Foundation Field Research Grant (2009) and Carroll D. Clark Research Award (2009). Data collected in 2007 includes 97 participants from Onquilha' (N=53) and Chimolon (N=44). Data collected in 2009/10 include 64 participants from Jocotán, Chiquimula.

3.2. LAB METHODS

3.2.1 DNA Extraction

DNA was extracted from buccal swabs, mouthwashes and whole blood using various methods depending on the time of collection and the state of the samples. DNA was first extracted from buccal cells using the QiaAmp® DNA Mini Kit (Qiagen®), a spin column based

extraction method, according to the manufacturer's instructions (Qiagen®, Valencia, CA). For samples with low DNA yield or where a second extraction was needed for additional analyses, samples were extracted using Evogen One. Evogen One™ is a proprietary salt, detergent, and heat method to lyse the cells. Lysed samples are centrifuged to pellet cellular debris, leaving supernatant containing PCR-ready DNA. Mouthwashes were extracted using Chelex, a resin that binds cellular debris after cells are lysed through heating. Like Evogen One™ the samples are centrifuged leaving the DNA suspended in supernatant. Whole blood was extracted using the Super Quick Gene® following manufacturer's protocol.

3.2.2 mtDNA Analysis

Mitochondrial DNA was analyzed using HVS1 sequencing and the Applied Biosystems SNaPShot assay for haplogroup assignment described by Nelson et al. (2007). The mtDNA HVS1 amplification, visualization, and purification were completed by the author in KU's Laboratory of Biological Anthropology. The ABI SNaPShot assay amplification, purification, minisequencing, and AFLP analyses were conducted by the author in the Department of Forensic Sciences Laboratory, George Washington University.

A fragment of the Control Region of the Hypervariable Sequence Region I (HVS1) of the mitochondrial DNA (mtDNA) molecule was analyzed for 52 samples, from nucleotide position (np) 15976 to 16422 using primers L-15976 (5'- CCA CCA TTA GCA CCC AAA GCT AAG -3'), H-16422 (5' - ATT GAT TTC ACG GGA GGA TGG - 3'), and for problematic samples, H-16498 (5' -

CCT GAA GTA GGA ACC AGA TG – 3'). The Polymerase Chain Reaction (PCR) reaction mix for the first primer pair consisted of 3.5 µl 5x buffer, 100 mM $MgCl_2$, 20mM dNTP, 1 U of Taq Polymerase, 25 pM of forward and reverse primers, 10-30 ng of DNA template, and molecular grade water to result in a 25 µl reaction mix per sample. For this reaction, the PCR amplification procedure consisted of an initial denaturation at 94°C for 1 minute, followed by 35 cycles of denaturation at 94° C for 40 seconds, annealing at 52° C for 30 seconds, and extension at 65° C for 2 to 4.5 minutes, and a final extension at 65° C for 5 minutes. For the second primer pair, the reaction mix consisted of 5 µL 5X buffer, 100mM $MgCl_2$, 1 µL purified BSA, 1 unit (U) Taq DNA Polymerase, 20mM dNTPs, 10pM forward and reverse primers, 1 ng DNA template, and molecular grade water to 25 µL volume. For this reaction, the initial denaturation was 95°C for 11 minutes; followed by 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds for 40 cycles; and a final extension of 72°C for 10 minutes. Amplification was verified by running amplicons out on a 1.5% Agarose gel and visualized using ethidium bromide on a UV illuminator. Amplicons were purified using a QIAquick® PCR Purification Kit (Qiagen, Valencia, CA) following the manufacturer's protocol, or ExoSAP DNAse. ExoSAP requires adding 1 µL of ExoSAP per 10 µL of PCR product and running in the thermocylcer at 37°C for 30 min, followed by 85°C for 15 min, then finishing with a soak or hold temp of 4°C. After purification, samples were sequenced using the aforementioned primers on an ABI PRISM 3130 (2007-2010) or 3730 (2011) in the University of Kansas DNA Sequencing Laboratory (Lawrence, KS).

HVS1 sequences were compared to the Cambridge Reference Sequence (CRS) to determine key diagnostic SNPs that could be used for mtDNA haplogroup assignment. These SNPs were then entered into the Genographic Project haplogroup predictor

(http://nnhgtool.nationalgeographic.com/classify/). To verify that these assignments were accurate, a subset of samples (N= 42) were typed using an ABI Systems SNaPShot assay developed by the George Washington University Department of Forensic Sciences DNA Laboratory. These tests were run by the author in the GWU laboratory. Each reaction consisted of four steps: 1) initial amplification of target regions for diagnostic SNPs, 2) purification of amplicons, 3) mini-sequencing using single base extension (SBE), and 4) Amplified Fragment Length Polymorphism (AFLP). The first step involves standard PCR amplification with a reaction mix containing 10.4 µL of GeneAmp PCR Gold reaction mix (Applied Biosystems), 0.5 µL of ampliTaq Gold DNA Polymerase (Applied Biosystems), 1 µL $MgCl_2$ (Applied Biosystems), 30.5 µL Primer Mix (GWU Forensic Sciences Laboratory), 6.6 µL ddH_2O, and 1 µL DNA template (~1 ng). A graduated cycling procedure was used for amplification with an initial denaturation of 95°C for 10 minutes; followed by 19 cycles of 95°C for 30 seconds, 50°C for 55 seconds, and 72°C for 30 seconds; then 19 cycles of 95°C for 30 seconds, 50°C +0.2°C per cycle for 55 seconds, and 72°C for 30 seconds; 11 cycles of 95°C for 30 seconds, 55°C for 55 seconds, and 72°C for 30 seconds; and final extension at 72°C for 7 minutes. These amplicons were then purified using 1.0 µL ExoSAP DNAse and 2.0 µL of amplified DNA and running the thermocylcer for a single cycle at 37°C for 70 minutes and 72°C for 20 minutes. The purified amplicons were then used for SBE mini-sequencing of the target SNP. The reaction mix consisted of 2 µL SNaPShot Ready Reaction Mix (Applied Biosystems), 7.8 µL mini-sequencing primer mix, and 3 µL purified PCR product. Finally, the new amplicons were prepared for AFLP by combining 10 µL of HiDi ladder and formadide mix to 1 µL of PCR product run on an ABI 3130. Data files were analyzed in GeneMapper software to obtain sizing results and determine SNP typing.

3.2.3. Y Chromosome Analysis

Y chromosome STRs and SNPs were characterized for male participants. Eight STRs (DYS19, DYS389I, DYS389II, DYS 390, DYS391, DYS 392, DYS 393, and DYS439) were amplified with fluorescently labeled primers in three separate PCR reactions by the author in the KU LBA following procedures outlined in Melton (2007). These amplicons were then sent to the KU Sequencing laboratory for Amplified Fragment Length Polymorphism (AFLP) analysis on an ABI 3730.

The first Y STR multiplex amplified DYS390, DYS391, and DYS393 and used the follow reaction mix: 4.4 µL 5X GoTaq Flexi Buffer; 95.0mM MgCl2; 40.0mM dNTPs; 0.4 µL purified BSA, 1.5 U Taq DNA Polymerase; 3.0 pM DYS390 forward (F) and reverse (R) primers, 2.5 pM DYS 391 F and R primers, and 2.0 pM of DYS393 F and R primers; ~40 ng DNA template; and molecular grade water to a 22 µL volume. Amplification reactions were run on an ABI 9600 with the following cycling procedure: initial melt at 94 °C for 3 minutes; followed by denaturation at 94°C for 30 seconds, annealing at 57°C for 30 seconds, and extension at 72 °C for 30 seconds for 40 cycles; a final extension at 72 °C for ten minutes; and then on hold at 4 °C.

The second Y STR multiplex amplified DYS19, DYS392, DYS438 and DYS439 and used the follow reaction mix: 3.6 µL 5X GoTaq Flexi Buffer; 77.5 mM MgCl2; 40.0mM dNTPs; 0.3 µL purified BSA, 1.5 U Taq DNA Polymerase; 3.0 pM DYS19 forward (F) and reverse (R) primers, 3.5 pM DYS 392 F and R primers, 1.0 pM of DYS438 F and R primers, and 1.0 pM of DYS439 F and R primers; ~40 ng DNA template; and molecular grade water to a 18 µL volume. Amplification

reactions were run on an ABI 9600 with the following cycling procedure: initial melt at 94 °C for 3 minutes; followed by denaturation at 94°C for 25 seconds, annealing at 54°C for 30 seconds, and extension at 72 °C for 30 seconds for 45 cycles; a final extension at 72 °C for three minutes; and then on hold at 4 °C.

Y STR DYS389 I/II was amplified in a singleplex reaction using the following protocols: 3.6 µL 5X GoTaq Flexi Buffer; 77.5 mM MgCl2; 16.0 mM dNTPs; 1 U GoTaq DNA polymerase; 0.3 µL purified BSA; 10 pM F and R primers, 40ng of DNA template, and ddH_2O to 18 µL volume. Amplification reactions were run on an ABI 9600 with the following cycling procedure: initial melt at 94 °C for 3 minutes; followed by denaturation at 94°C for 25 seconds, annealing at 54°C for 30 seconds, and extension at 72 °C for 30 seconds for 45 cycles; a final extension at 72 °C for three minutes; and then on hold at 4 °C.

Each of the three PCR reactions were combined in a single tube and diluted 1:100 before AFLP analysis on an ABI 3730. AFLP was completed at the KU Sequencing laboratory by Dr. Mike Grose. The standard ABI LIZ(500) ladder was used for sizing of fragments and a DNA control of known repeat length for each marker was used for determining allele size. Peak Scanner Software v1.0 (Applied Biosystems) was used to analyze data and type individuals.

A novel probe technology was used for Y chromosome SNP analysis. The technique involves a single primer pair to amplify the region of interest, normal PCR reagents, and Hybeacons® Probes in a single PCR reaction. HyBeacons® are high definition fluorescently labeled PCR probes with two fluorophores. When a probe binds with the target DNA sequence, the level of fluorescence from the probe intensifies, and is measured through melt curve analysis. The difference among melt curves allows the researcher to differentiate among the

wild type and the polymorphic SNPs. Thus, genotyping requires only one reaction per SNP. This new method was used applied to four common Y chromosome SNPs used for haplogroup assignment (Q-M242, Q-M3, R-M269, Q-P36.2) in Native American populations.

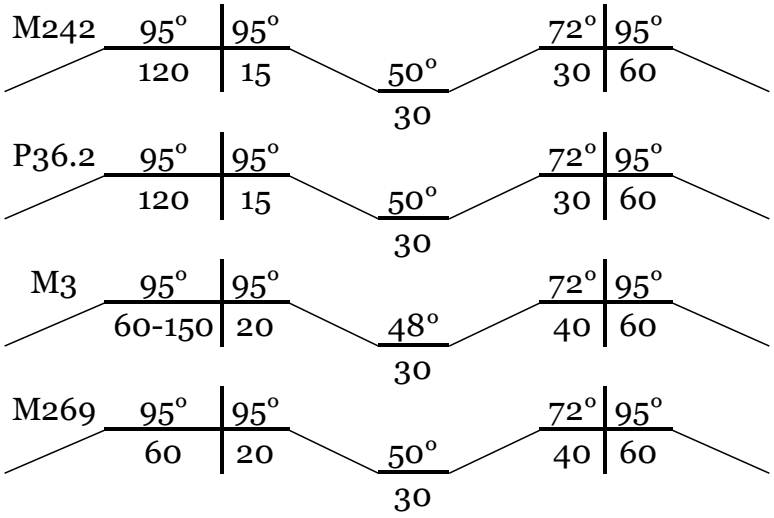

Figure 3. 3. The cycling protocols for each of the four Y SNPs analyzed with Hybeacons^{TM} Probes.

Table 3. 1. Sequences for the primer pairs and probes. Target SNPs are highlighted in bold within the Probe sequences, with FLD marking the fluorescently-labeled nucleotide.

Y-SNP	*Haplogroup*	*Mutation*	*Primers*	*Probes*
M242	Q	C > T	(For) AACTCTTGATAAACCGTGCTG	CAAGG[FLD-T]GCTCT[FLD-T]CGCATTGGTC-3'
			(Rev) TCCAATCTCAATTCATGCCTC	
P36.2	Q1	G > T	(For) TGAAGGACAGTAAGTACACA	GGATAA[FLD-T]GGATATA[FLD-T]ATAGAGAGAGAAT-3'
			(Rev) TAAGTCCATTGATCTACAGA	
M3	Q1a3a	C > T	(For) TAATCAGTCTCCTCCCAGCA	CTGGGAC[FLD-T]GATAAT[FLD-T]AGGAAGAGC-3'
			(Rev) AAAATTGTGAATCTGAAATTTAAGG	
M269	R1b	C > T	(For) CTAAAGATCAGAGTATCTCCCTTTG	GGTTTGGT[FLD-T]AATCC[FLD-T]GGTAAATTG-3'
			(Rev) AAATTGTTTTCAATTTACCAG	

Hybeacon Probes were designed by the author and Evogen™ and synthesized by Sigma Aldrich using Hex fluorophores. All probe assays were run on a Rotor Gene Q by Qiagen. Each reaction consisted of the following reaction mix: 12.5 µL Promega Master Mix, 2.5 pM non-target strand primer, 25 pM target strand primer, 6.0 pM HyBeacons Probe, ~20 ng of template DNA, 5.75 µl dd H_2O. Figure 3.3 provides the cycling profiles for each probe. Cycles were added to buccal sample amplification in five cycle increments in order to achieve adequate amplification. Final cycle number ranged from 45-60 cycles. Also, annealing temperature was decreased as much as 2°C to increase amplification for samples extracted from buccal swabs. Melting profiles for each probe are shown in Table 3.2. All fluorescent data were acquired during the annealing phase and continually during melt curve analysis. SNP typing was confirmed on a subset of samples by direct sequencing on ABI prism 3130.

Table 3. 2. Melting protocols for each of the four Y SNPs analyzed with Hybeacons™ Probes.

Marker	Temp. range	Temp. Increment	Time Increment
M242	45°C - 80°C	0.3°C	1 sec.
P36.2	45°C - 75°C	0.3°C	1 sec.
M3	45°C - 65°C	0.3°C	1 sec.
M269	40°C - 70°C	0.3°C	1 sec.

For the remaining unknown haplotypes, Y-STR profiles were entered into Whit Athey's Haplogroup Predictor (http://www.hprg.com/hapest5/index.html) for haplogroup assignment. STRs (DYS19, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439) were entered and prior probabilities of European location were left equal. While this on-line identification tool is designed to recognize European Y chromosome haplogroups, it includes some African and Native American haplogroups (E and Q respectively). Also, non-European haplogroups are suggested by low fitness and probability scores for the major European haplogroups included in the predictor (G, H, I, J, L, N, Q, R, T, and associated sub-haplogroups).

3.3 ANALYTICAL METHODS

3.3.1 Within Population Variation

To determine the amount of variation within populations, nucleotide diversity was measured using mtDNA sequence data (Nei 1987; Nei and Li 1979; Tajima 1983). Also, gene diversity was calculated for mtDNA and Y chromosome STR haplotype data using Nei's (1987) method, which is less affected than is nucleotide diversity by recent evolutionary events and stochastic changes in allele frequency Both of the tests are equivalent to estimating average heterozygosity in a population, but for haploid data like Y chromosome and mtDNA. Nucleotide diversity (or average gene diversity for STR data) was calculated as

$$\hat{\pi}_n = \frac{\sum_{i=1}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij}}{L},$$

(Equation 3. 1)

where p_i is the probability of the i[th] sequence in the population, p_j is the probability of the j[th] sequence in the population, $\hat{d}_{ij}$ is the number of nucleotide differences (or mutational differences in the case of STRs) between the i[th] and the j[th] sequence, and *L* is the number of loci. For Y STR data, the mean pairwise differences are reported, which are equivalent to the numerator of the above equation multiplied by *n/n*-1, where *n* is equal to the sample number. Gene diversity was calculated as

$$\hat{H} = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right),$$

(Equation 3. 2)

, where *n* is the sample size, *k* is the number of haplotypes, and p_i *is* the frequency of the *i*[th] haplotype. These tests were performed using Arlequin 3.5.1.2 (Excoffier and Lischer 2010a).

3.3.2 Variation Among Populations

In order to characterize the amount of variation among the populations studied, several methods of analysis were chosen including Analysis of Molecular Variance (AMOVA), Median-Joining Network Analysis (MJ), Multi-dimensional Scaling Plots (MDS), and a Neighbor-Joining

Tree (NJT). Each of the methods is designed to elucidate the relationship among populations by either partitioning the variation, co-variation, or mutational differences among individuals and groups.

3.3.2.1 AMOVA

AMOVA is an extension of a basic Analysis of Variance (ANOVA) that incorporates the mutational differences among haplotypes to calculate the observed variance among groups. The analysis begins by the investigator defining the various groups to be analyzed from the populations in the study. For the purposes of this project several models were tested, first by analyzing the mtDNA variation within and among Maya populations and arranging them based on major Maya Region (Ch'orti' = Central, Yucatec = Northern, and Poqomchi' and K'iche = Southern). Additionally, populations were grouped into major geographical/cultural regions (Mesoamerica including Mexico, Guatemala, Honduras, and El Salvador; Central America including Nicaragua, Costa Rica, and Panama; and South America) for both mtDNA and Y chromosome data. Finally, to investigate the relationship between linguistic and genetic diversity, groups were arranged by major language family according to Campbell's classification of American Indian languages (Barbacoan, Chibchan, Chocoan, Ciboney, Jêan, Mapudungu, Mapurean North and South, Maya, Movima, Otomanguen, Quechua, Tarascan, Tupían, Uto-Aztecan, Yanomaman, and Yuracare) (Campbell 1997). All populations where the language or language family could not be identified were removed.

These models of genetic structure are tested with AMOVA by using the covariance among haplotypic distances and computing hierarchical fixation indices similar to Wright's (Wright 1951; Wright 1965) F_{st} values. Kimura (Kimura 1980) two parameter distances with a gamma correction of 0.26 (Excoffier and Yang 1999; Meyer et al. 1999) were used for mtDNA sequence data and Slatkin's (Slatkin 1995) Linearized F_{st} (R_{st}) for Y chromosome STR data. Kimura 2p distance weighs transversions and transitions differently in the calculation to account for the difference in mutation rate between the two. The 2p distance is calculated as,

$$\delta_{jk}^2 = \frac{1}{2}\ln(1 - 2\hat{P} - \hat{Q}) - \frac{1}{4}\ln(1 - 2\hat{Q}),$$

(Equation 3. 3)

where $\hat{P}$ is the frequency of transitions and $\hat{Q}$ is the frequency of transversions between the sequences. Slatkin's R_{ST} estimates the sum of squared differences between repeat numbers of two haplotypes to measure distance between two Y STR haplotypes as,

$$\hat{d}_{xy} = \sum_{i=1}^{L}(a_{xi} - a_{yi})^2,$$

(Equation 3. 4)

summing the differences between the x^{th} and y^{th} population at the i^{th} locus. The AMOVA, like an ANOVA, compares the proportion of the Sum of Squared Differences (SSD) to the Mean Square Error (MSE) among the hierarchical groups. To illustrate, the Total SSD is calculated as,

$$SSD(T) = \frac{1}{2}N \sum_{j=1}^{2N} \sum_{k=1}^{2N} \delta_{jk}^2 \ ,$$

(Equation 3. 5)

where *N* is the number of haplotypes. MSE is equivalent to the covariance of the differences among the haplotypes within each level or the SSD multiplied by the appropriate number of degrees of freedom. The F ratios for AMOVA are referred to as ϕ statistics and are calculated as

$$\Phi_{ST} = \frac{\sigma_a^2+\sigma_b^2}{\sigma_T^2} \ , \ \Phi_{SC} = \frac{\sigma_b^2}{\sigma_b^2+\sigma_c^2} \ , and \ \Phi_{CT} = \frac{\sigma_a^2}{\sigma_T^2} \ ,$$

(Equation 3. 6)

where σ_a^2 is the covariance among groups, σ_b^2 is the covariance among populations within a group, σ_c^2 is the differences among individuals in a population, and σ_T^2 is the total covariance of haplotypes (Excoffier and Lischer 2010b; Excoffier et al. 1992). Each of these ratios describes the proportion of the total variation explained at that level of grouping. A significance value is provided by creating a null model through permutation of the haplotypes among populations and groups and comparing the observed model to the null model. This comparison may result in a negative value for variance, unlike in normal tests of ANOVA. All AMOVAs were performed using Arlequin 3.5.1.2 (Excoffier and Lischer 2010a).

3.2.2.2 Median Joining Network Analysis

In order to examine the phylogenetic relationship between the focus population and closely related comparative populations and highlight underlying genetic structure of the Maya, median-joining (MJ) network analysis (Bandelt et al. 1999) was performed for each of the four major Amerindian mtDNA haplogroups using sequence data and the Native American Y haplogroup Q for the two Maya populations using STR data. Network analyses such as MJ are often better suited for visualizing the evolutionary relationship between populations with small genetic distances, and MJ was specifically designed to deal with multistate data (non-binary sequence data) such as mtDNA sequence data, as well as genetic systems subject to homoplasy, such as STRs (Bandelt et al. 1999; Bandelt et al. 1995). Also, MJ provides alternative evolutionary branches by focusing on the evolutionary history of a single haplogroup and highlighting similarities between populations that may be due to homoplasy as reticulations. MJ works in two phases, the first of which involves selecting the median vectors for all haplotypes by creating a distance matrix of all haplotypes among the populations. The second phase is the construction of the network through the calculation of the minimum spanning tree, or the tree that generates the shortest distance between all points and minimizes midpoints. Once ancestral nodes have been identified, they can be used to determine the average distance to the node. All network analyses were carried out using Network ver. 4.0 (www.fluxus-engineering.com).

3.2.2.3 Multi-dimensional Scaling Plots

Ordination techniques are employed to create a graphical display of differences or similarities in multivariate data into reduced dimensional space to aid in interpreting relationships among variables. Distance matrices (Kimura 2p for mtDNA and Slatkins R_{st}, described above) were used to create a Multi-dimensional Scaling (MDS) plot. This analysis decreases the variation among groups to as few axes as possible to more easily visualize multi-dimensional distances among populations. This is achieved by taking the original distance matrix and comparing it to randomly generated distance matrices created using the specified number of planes. A regression is then run between the resulting distance matrix and the original distance matrix to assess the fit. There are several tests of goodness of fit available. For the purposes of this study, Kruskal's stress formula one and two are utilized. As is implied by the term, the "stress" actually informs the researcher of the poorness of the fit between the two matrices, or the magnitude of the disparities (the error or deviations from the original dissimilarity matrix). The closer the stress value is to zero, the better the fit between the MDS plot and the original distance matrix. Any number of permutations, within the parameters of the original matrix, can be completed until the fit has been maximized (stress has been minimized) (Manly 2005; Sturrock and Rocha 2000). All MDS plots were performed in NTSYSpc version 2.1 (Rohlf 2000).

3.2.2.4 Neighbor-Joining Tree

A Neighbor Joining Tree (NJT) was constructed using the Kimura 2p distance for mtDNA sequence data and Slatkin's linearized distance for Y chromosome STR data. The purpose of a NJT is to create a topology that minimizes the total branch length for the tree or represents the tree with the shortest evolutionary time based on the given distance matrix. Again, the purpose of NJT is to visualize the relationship between these populations, but by looking at the most likely closest "neighbor", or closest operational taxonomic unit (OTU) (Saitou and Nei 1987). This OTU can supply a means to infer not just overall relatedness, but possible phylogenetic relationship. As with any transformation of data, some information is lost when the distance matrix is transformed into the NJT. To assess the fit of the tree to the original data, a cophenetic distance matrix is created from the tree and then compared to the original distance matrix using a Mantel randomization test. The NJT, cophenetic matrix, and Mantel randomization test were performed using NTSYSpc version 2.1 (Rohlf 2000).

3.3.3 Measures of Forces of Evolution

While many of the tests above highlight differences among populations, they do not explicitly test for the effects of forces of evolution and demographic changes in population size. So to fully understand what causes the differences within and among populations, several

methods were chosen to evaluate the effects of evolution and demographic changes including Tajima's D, Fu's F_s, Mismatch Analysis, and comparison of r_{ii} and diversity.

3.3.3.1 Neutrality Tests

Two tests of neutrality (D and F) were performed on the mtDNA sequence data to elucidate any possible effects of natural selection and/or fluctuations in population size. Tajima's D (Tajima 1989) and Fu's F_s (Fu 1997) are both based on an infinite-site model without recombination making them applicable to haploid data such as mtDNA. However, Fu's F uses haplotypic distributions while Tajima's D uses pairwise sequence data and the assumption of a constant mutation rate to test the neutral model, and is calculated as,

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_S}{\sqrt{Var(\hat{\theta}_\pi - \hat{\theta}_S)}},$$

(Equation 3. 7)

where $\hat{\theta}_\pi$ is equal to the average number of nucleotide differences in a population and $\hat{\theta}_S$ is equal to the number of segregating sites. So, Tajima's D compares the proportion of average number of pairwise differences $\hat{\theta}_\pi$ to the total number of nucleotide differences $\hat{\theta}_S$. During a population expansion, we expect to see an increase in the total number of nucleotide differences, which will make $\hat{\theta}_S$ larger. If $\hat{\theta}_\pi$ is larger, this means that there are fragmented and deep internal branches indicating genetic drift. Therefore, stable populations are expected to approach zero.

For Fu's Fs, we are testing for the same evolutionary effects, but this test usually proves more sensitive to population expansion rather than genetic drift. F_s is equal to the natural log of the probability of observing more haplotypes than are present in the data (*S*') divided by the probability of not being present (incorporating the inverse probability), calculated as

$$F_S = \ln\left(\frac{S\prime}{1-S\prime}\right) ,$$

(Equation 3. 8)

so that negative values result from having higher than expected number of haplotypes in the population and positive values are lower than expected. Therefore, like with Tajima's D, large negative values are indicative of a population expansion and large positive values are the result of genetic drift. Both tests were performed using Arlequin 3.5.1.2 (Excoffier and Lischer 2010a).

3.3.3.2 Mismatch Analysis

Mitochondrial DNA sequences were used to conduct a Mismatch analysis (Rogers and Harpending 1992), which produces a distribution of pairwise differences between individuals within a population. This distribution may provide evidence for population expansion, stability, or bottlenecks. Assuming an infinite sites model with no recombination, a population in drift-mutation equilibrium will display a unimodal mismatch distribution with a peak at zero mismatches. For populations which have undergone a significant population expansion, this distribution will be unimodal with a peak greater than zero mismatches. Alternatively, a unimodal distribution may be indicative of directional selection. When a bimodal distribution of

mismatches is present, then it can be indicative of both current population stability (with a peak at zero mismatches) and a remnant past population expansion (if the greater than a zero peak is significant) or, alternatively, selection and population bottlenecks can produce bimodal distributions with peaks greater than zero. Additional analyses must be conducted to delineate among these hypotheses (Rogers and Jorde 1996; Schneider and Excoffier 1999). For human mtDNA hypervariable sequence data, selection is considered unlikely, but a population bottleneck can be revealed in combination with a network analysis with deep lineages in a star-like cluster (Ramirez-Soriano and Nielsen 2009). In this dissertation, mismatch distribution were created using Network ver. 4.0 (www.fluxus-engineering.com).

3.3.3.3 Comparison of r_{ii} vs. Diversity

Finally, to examine the effects of gene flow and genetic drift on these populations, distance from the centroid (r_{ii}) was plotted against gene diversity for both mtDNA sequence and haplogroup data and Y chromosome STR and haplogroup data. This comparison is similar to Harpending and Ward's (1982) method of examining a regression between distance from the centroid and heterozygosity. This relationship should be linear, assuming isolation by distance without any additional gene flow or genetic drift. Any populations plotting above the theoretical regression line may be undergoing significant gene flow, while those populations below the regression line may be experiencing genetic drift. Distance from the centroid is calculated using haplogroup frequencies, where

$$r_{ii} = \sum_{i=1}^{n} \frac{(p_i - \bar{p})^2}{\bar{p}\,(1-\bar{p})},$$

(Equation 3. 9)

and p_i is equal to the haplogroup frequency in the i^{th} population. Harpending and Ward (1982) acknowledged that there is a linear relationship between heterozygosity and r_{ii} *when* no forces of evolution are operating on the populations. Consequently, any observed deviations from the theoretical regression line are indicative of a population experiencing either gene flow or genetic drift, with those populations resting above the theoretical line (expressing more variation than expected) are likely experiencing gene flow, while those populations below the line (having less diversity than expected) are likely under the influence of genetic drift (Harpending and Ward 1982). Again, directional selection may also cause a population to express less than expected variation; however, this is unlikely for mtDNA HVS1 data.

3.3.4 Phylogeographic Methods

Phylogeography is a relatively young subfield of systematics, first established in 1987, the basic theories of which were developed within a few decades prior with its roots in biogeography (Avise et al. 1987) and gene geography (Manni et al. 2004). Phylogeography combines genetic data and geographic dispersion of populations to interpret the biological relationship of species and extends these interpretations into a temporal and ecological framework. The relationship between changes in gene frequencies and geographic space can

be used to infer micro (traditionally the focus of population genetics) and macro (phylogenetics) evolutionary events operating on species and allow researchers to determine the phylogenetic relationships binding their history (Avise 1998; Avise 2000; Avise et al. 1987). Advances in genomic technology and statistical theory have led to a wide array of applications for phylogeographics, including species conservation, evolutionary ecology, human evolution, comparative phylogenetics and species co-evolution (Avise 1998). While the usual applications of these methods utilize dispersion in geographical space, due to the nature of these methodologies, any distance parameters that can be estimated and observed can be entered into a model (i.e., geographic, linguistic, cultural, temporal distance). Therefore, phylogeographic methods are applied to both spatial data and linguistic data in some of the following procedures.

3.3.4.1 Mantel Randomization

In order to examine the relationship of phenotypic variation and geography, a Mantel randomization test was run between the standardized genetic distance matrices and the standardized geographic distance matrix for both mtDNA sequence data and Y chromosome STR data. A Mantel test involves holding one matrix constant and creating a randomly configured matrix from the second, then creating a correlation statistic for the constant matrix and the randomized matrix. A distribution of the correlation statistics is created from the randomization tests performed. The correlation between the two original distance matrices is then compared to the randomized correlation distribution to see if the results are significant. If

there is no true relationship of the two matrices, then one would expect that the correlation between the two distance matrices would be similar to the correlation among any of the randomly created matrices and the constant matrix. However, if there is an underlying relationship, then one would find that the correlation is significantly positive or negative when compared to the average random correlation. Mantel tests were performed using NTSYSpc version 2.1 (Rohlf 2000). This test was run with 1000 randomization attempts. Since there are *k*! possible random matrices that can be generated from the original distance matrices for *k* populations, the numbers were decreased to lower the computation time and fit within the limits of the computer program. Geographic spheroidal distance matrices were created using Geographic Distance Matrix Generator version 1.2.3 (Ersts 2011).

Mantel tests were also performed comparing linguistic distance and genetic distance. Linguistic distance was calculated by first using Campbell's (1997) grouping of Native American Languages in a hierarchical fashion. Distance was estimated by including direct linguistic neighbors as having a distance of 1. Distances were then estimated in clusters, with all members of the next hierarchical grouping having a distance of one greater than the lower level until the largest grouping that Campbell reported confidence in (major language areas) were reached. At this point the highest clustering level of all groups was used for estimating

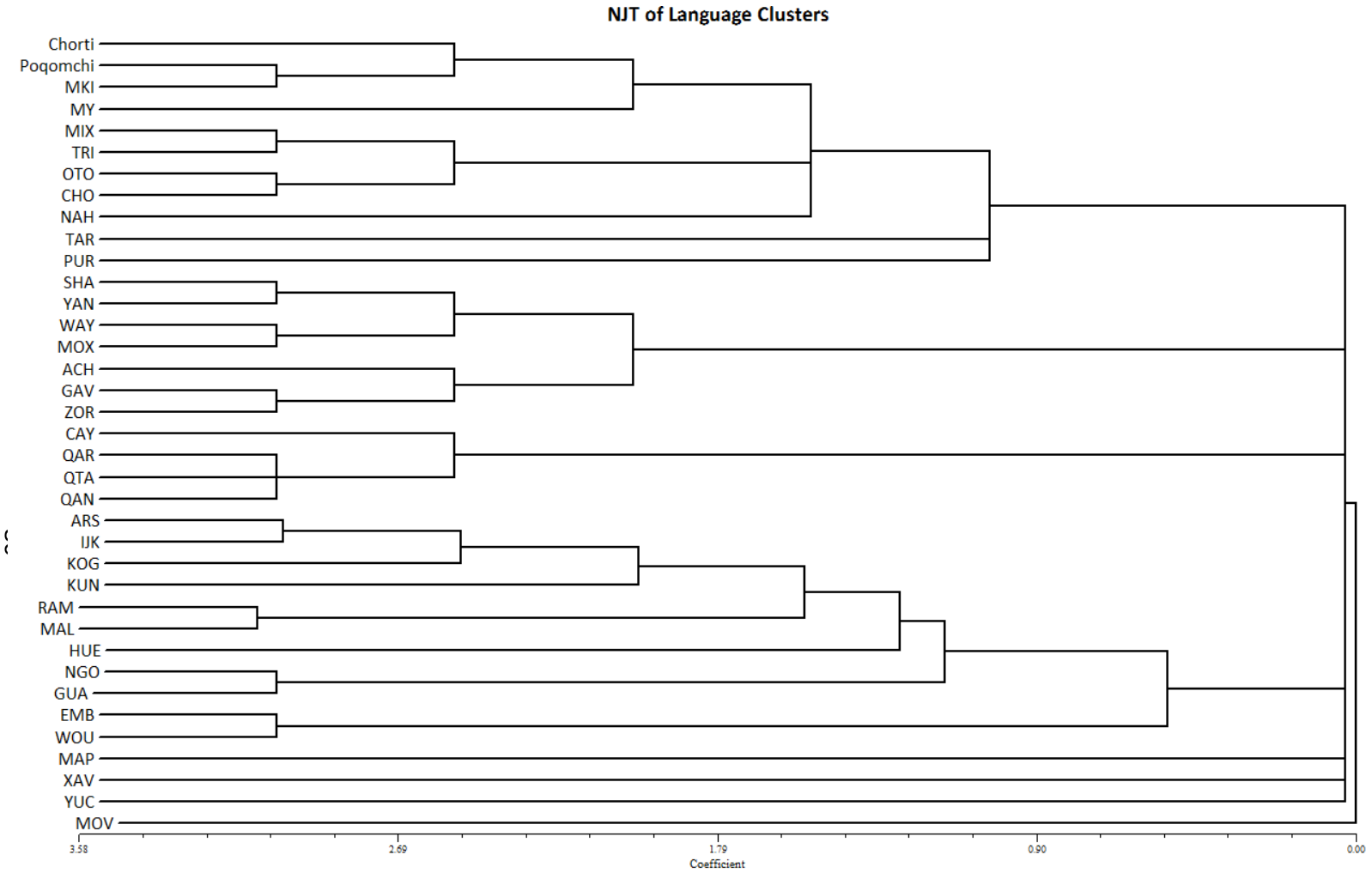

Figure 3. 4. The hierarchical typology used to represent the linguistic relationship among the populations compared in the study.

the cluster distance among language groups. This method ignores branch length, so assumes equal time (distance) since separation of major language groups. These distances were then used to generate a NJT tree to confirm language typology (Figure 3.4) and then standardized for comparison with the standardized genetic distance matrix.

3.3.4.2 SAMOVA

Since limited gene flow leads to population differentiation, a goal of phylogeography is to identify regions that may be influenced by barriers to gene flow. Researchers often assume that greater geographic distance leads to greater genetic distance; however, the relationship between gene flow and geography is not always that simple, especially among human groups. Because of these challenges, several methods have been used to identify barriers to gene flow within geographic space. One such method, Spatial Analysis of Molecular Variance (SAMOVA) (Dupanloup et al. 2002), identifies groups of populations that maximize the variation among the groups while increasing the homogeneity within the newly defined groups. To identify the groups of populations, a program, SAMOVA (named so after the method), was developed that runs a simulated annealing algorithm to select populations geographically adjacent to one another to form a new group. The algorithm is repeated until an optimal value of F_{ct} is achieved - in this case when the variation within the groups is minimized and variation among the groups is maximized. F_{ct} is calculated in the same manner as described for AMOVA.

One of the benefits of using SAMOVA is that it requires no prior knowledge of potential group structure, including group division or group number. The software package allows the user to choose anywhere from two to twenty groups, for this study groups were ordered into two to five groups and the model with the highest F_{ct} is presented. Additionally, the program requires only two input files, one containing the raw geographic coordinates and another containing the genetic data. While the program identifies those groups that maximize the variation among the populations, simulations have shown that the program does not always identify true genetic barriers (Dupanloup et al. 2002). Since SAMOVA always regroups to maximize estimates of F_{ct}, the final formulation of populations may not include populations that are geographically adjacent to one another. Furthermore, SAMOVA was shown to perform with less accuracy when compared to Monmonier's Maximum Difference Algorithm (see discussion below) when true genetic barriers are known and becomes less reliant with fewer markers. SAMOVA was performed on both mtDNA sequence data and Y chromosome STR data in SAMOVA 1.0 (Dupanloup et al. 2002).

3.3.4.3 Monmonier's Maximum Difference Algorithm

Monmonier's Maximum Difference Algorithm was proposed in 1973 as a computationally simple method of finding potential discontinuity and barriers of contiguous change in variables across space. Mark Monmonier, a cartographer/geographer, used demographic data, such as birth rates in the U.S., to test the validity of assigned borders in

maps to most appropriately display data. Monmonier began his analysis by calculating distance between the characteristics within the defined units of space (e.g., states in Midwestern U.S.A.). The algorithm begins by finding the border along the edge that specifies the greatest distance between two points, or the greatest change in the characteristic of interest. The borders along the next polygon are examined again for the greatest difference, and the barrier continues. The process is repeated over and over again until the edge of the space is reached or the barrier itself is reached (Monmonier 1973). The algorithm may be run any number of times until the target number of groups or barriers have been reached. In subsequent runs of the algorithm, the barrier may begin along any barrier within the space and run in opposite directions until the edge or first barrier is reached at either end (See Figure 3.5 for a basic representation of resulting barriers in a theoretical polygon). These barriers represent the space of most rapid genetic change and, therefore, possible barriers to gene flow between adjacent populations, or at least barriers between populations sharing the smallest amount of gene flow among those sampled.

Monmonier's algorithm was employed to show any possible genetic barriers among the populations compared in this study. Monmonier's algorithm, like other phylogeographic methods, identifies geographical areas with potential genetic barriers to gene flow by locating areas of rapid change in gene frequency. This method was chosen over others as it has been shown effective in locating the correct genetic barriers when only using one marker, such as mtDNA sequence data (Dupanloup et al. 2002). The computer program, Alleles in Space (Miller 2005) was used to apply Monmonier's algorithm to the sequence and STR data.

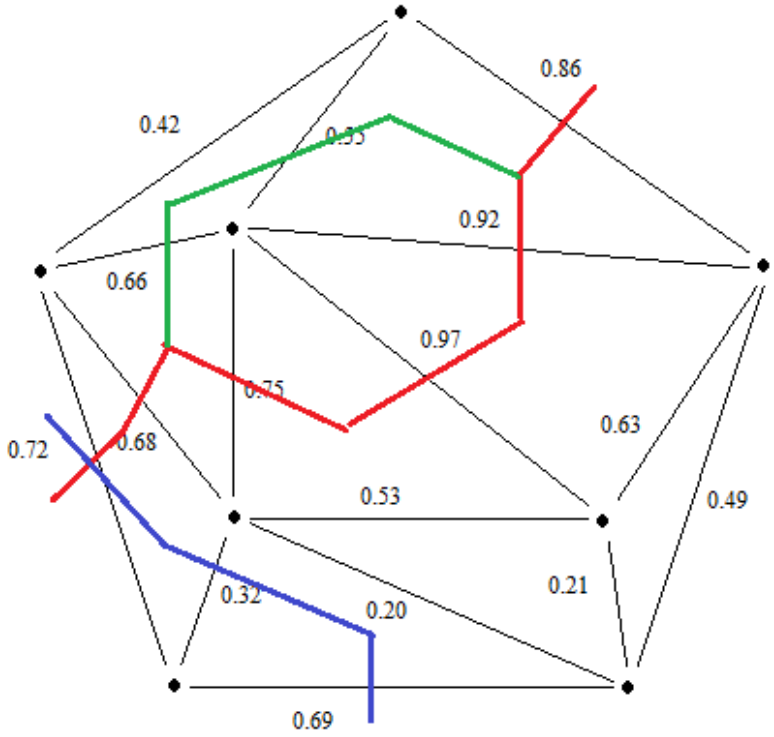

Figure 3. 5. Theoretical results of Monmonier's Maximum Difference Algorithm. Each black dot represents a potential sample location in theoretical space, with black lines representing the shortest distance between these points, as in the Delaunay Triangulation. The red line depicts the first barrier detected by beginning along the edge of the space and across the midpoint of the border with the greatest rate of change, or the greatest difference, and then running across subsequent borders lying adjacent to the first barrier and possessing the highest distance. The blue line represents the second selected barrier and the green the third.

Alleles in Space (AIS) is statistical genetics software that runs phylogeographic analyses, including Monmonier's Maximum Difference Algorith (Miller 2005). The user needs two input files for AIS, a file containing the original genetic data (e.g. sequence data and diploid or haploid genotypic data) and a file containing spatial data (X/Y coordinates, longitude and latitude). AIS calculates genetic distance using the following equation

$$D_{ij} = \sum_{k=1}^{n} \frac{d_k}{n}$$

(Equation 3. 10)

This equation is similar to Nei's genetic distance but for differences between individuals rather than populations. Equation 10 is applied to both haplotypic and sequence data where d_k is either 0, if individuals i and j have the same allele (or nucleotide if it is sequence data) at locus k or 1 if the allele is different (Miller 2005). This distance can then be regressed against geographic distance and the residual distances are then used for the Monmonier's algorithm. The program follows by projecting the sample localities into two-dimensional space using Delaunay Triangulation (see Figure 3.5) (selecting the shortest distance between two centroids, creating polygons in which all points are closest to their own centroid and not present in two polygons). The computer program then initiates Monmonier's Maximum Difference Algorithm and runs until the specified number of barriers have been detected. The resulting polygon and barriers are then displayed. These barriers represent the space of most rapid genetic change and, therefore, possible barriers to gene flow between adjacent populations, or at least barriers between populations sharing the smallest amount of gene flow among those sampled.

3.3.4.4. Interpolated Genetic Landscape

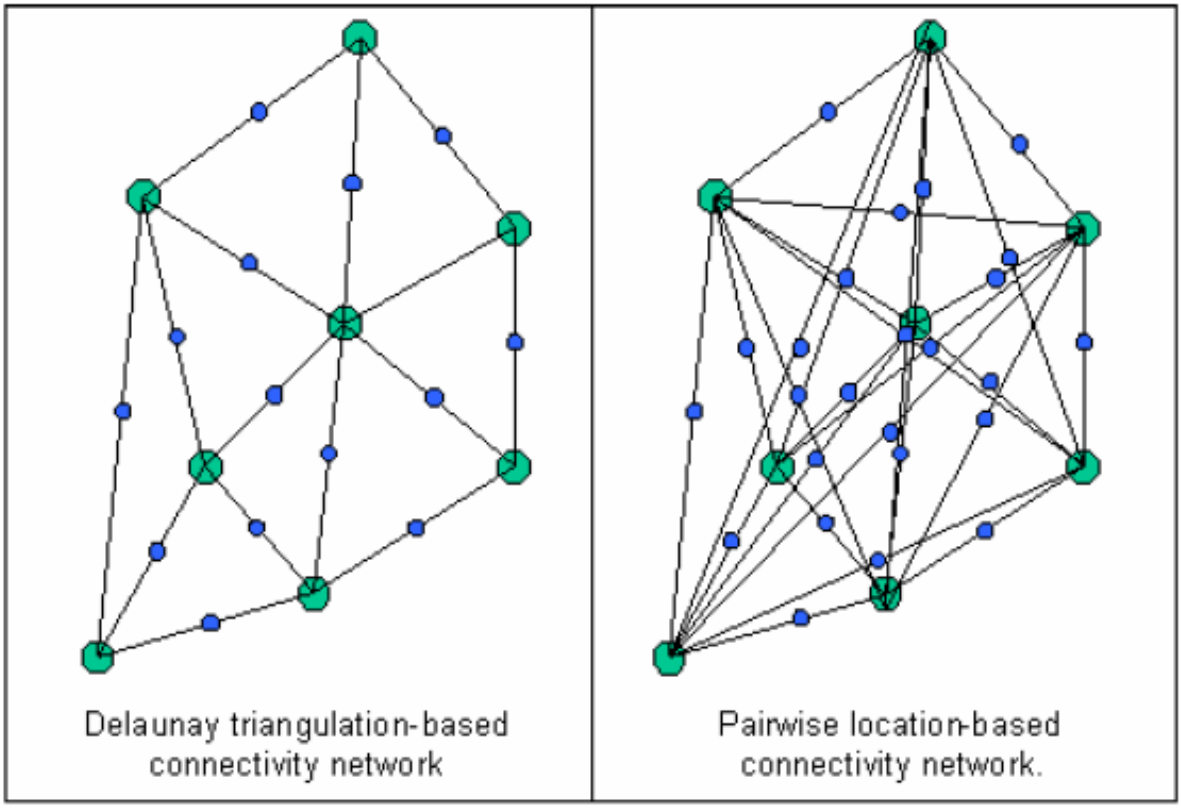

Figure 3. 6. Graphical representation of options for connectivity networks used in manufacturing the Interpolated Genetic Landscape (Miller 2005).

Interpolated Genetic Landscape is another means by which investigators can visualize changes in genetic distance across geographic space (Miller 2005). Similar to Monmonier's Maximum Difference, Interpolated Genetic Landscape begins by creating a two-dimensional connectivity network of the geographic coordinates. The user has two options for the creation of this network. The first is the same as for Monmonier's: using Delaunay Triangulation to connect all sample locations (marks as green circles in Figure 3.6). The other option available is the pairwise location-based connectivity network, which creates a denser network connecting

each collection site to every other one. In human population studies often the collection site for an entire population is recorded and humans tend to live in towns, so the dispersal of members of a sample may not be as continuous. Therefore, the second option is preferable when fewer collection sites are in the network, so this was selected for the current study. The next step assigns a genetic distance to the midline of each vector in the connectivity network (marked by blue circles in the picture). Next, a grid is created with a user-specified density and a third dimension is added to the plot, so that X and Y coordinates are points within the original connectivity network and Z coordinates are weighted inverse distances calculated as,

$$z = \frac{\sum_{i=1}^{n} w_i \times Z_i}{\sum_{i=1}^{n} w_i}$$

(Equation 3. 11)

where Z_i is the genetic distance along the nearest vector,

$$w_i = [(X_i - x)^2 + (Y_i - y)^2]^{\frac{a}{2}} \quad when\ X_i <> x, Y_i <> y$$

$$w_i = 1 \quad when\ X_i = x, Y_i = y$$

(Equation 3. 12)

and *a* is the value by which the genetic distances are weighted and vary between 1 and 0. When *a* = 1 all points within the grid are weighted equally, but as *a* approaches 0, values closest to the X,Y coordinate in the grid are weighted more heavily. The program then provides a

graphical output with only those X, Y, and Z coordinates that fall within the original polygon of data collection sites. Any peaks above the horizontal plane ($Z > 0$) can be interpreted as areas of greater genetic divergence, and peaks that fall below the plane ($Z < 0$) are areas of greater similarity (Miller 2005). Again, the residual distances were chosen in the case to create the genetic landscape.

A benefit of Interpolated Genetic Landscape is the possibility of qualitative inference of the data. Both SAMOVA and Monmonier's Maximum Difference Algorithm provide a quantitative approach to choosing groups of highly differentiated populations. Interpolated Genetic Landscapes do not choose groups or identify barriers between populations, but present a graphical representation of distances across space allowing the researcher to form their own interpretation (Miller 2005). However, the qualitative nature of the plot may lead to inconsistency in data interpretation making conclusions less reliable unless paralleled using other methods of analysis.

CHAPTER 4. RESULTS

This chapter presents the results of the mtDNA haplogroup and sequencing and Y chromosome haplogroup and STR analyses. Analytical methods described in the previous chapter were applied to new data acquired from the Ch'orti-speaking and Poqomchi'-speaking Maya of Guatemala and compared to other Mesoamerican, Central American, and South American native populations. These analytical methods include gene and haplotype diversity (π, *h*), neutrality test statistics (D and F_S), phylogenetic analyses (e.g. MJ network and NJT), AMOVA, and phylogeographic methods (e.g., Mantel tests, SAMOVA, Monmonier's Maximum Difference Algorithm, interpolated genetic landscape), among other methods described in chapter 3. The following sections will first present results on mtDNA analyses for each of these categories of inference, followed by the results from Y chromosome analyses.

4.1. MITOCHONDRIAL DNA

4.1.1. Haplogroup and Haplotype Results

The mitochondrial DNA results for haplogroup assignment in the Poqomchi' and Ch'orti populations are both similar to one another and to other Central American and Mesoamerican native populations (Table 4.1). Like nearby populations, the Poqomchi' and Ch'orti both exhibit high frequencies of haplogroup A (82% and 70%, respectively) and moderate frequencies of haplogroup C (12% and 25%). Both haplogroups B and D are completely absent from the

Ch'orti' speaking Maya, while only haplogroup D is absent from the Poqomchi'. The Ch'orti' Maya exhibit both African and European admixture, although at almost negligible frequencies (3.5% and 1.5% respectively). However, there is no evidence of maternal non-native admixture present in the Poqomchi' Maya.

Table 4. 1. Haplogroup frequencies for mtDNA used for calculating distance from the centroid. The haplotype diversity used as comparison to distance to the centroid is included in this table, as well.

Code	Population	N	A	B	C	D	Other	r_{ii}	*h*
ACH	Aché[22]	63	10%	90%	0%	0%	0%	2.081	0.204
ARS	Arsario[16]	28	68%	0%	32%	0%	0%	1.624	0.725
CAY	Cayapa[19]	30	30%	40%	10%	20%	0%	0.164	0.837
CHO	Chorotega[14]	30	73%	10%	0%	17%	0%	1.736	0.670
CIB	Ciboney[12]	15	7%	0%	60%	33%	0%	1.669	0.943
ELS	El Salvador Mixed[20]	90	91%	2%	2%	0%	4%	3.295	0.919
EMB	Emberá[9]	44	23%	52%	25%	0%	0%	0.340	0.942
GAV	Gavãio[23]	28	14%	14%	0%	71%	0%	3.091	0.862
GUA	Guayami[14]	50	78%	22%	0%	0%	0%	1.956	0.819
HUE	Huetar[14, 15]	67	64%	16%	0%	15%	4%	1.435	0.787
IJK	Ijka [16]	31	90%	3%	6%	0%	0%	2.903	0.185
KOG	Kogi [16]	21	67%	0%	33%	0%	0%	1.584	0.524
KUN	Kuna [2]	63	71%	29%	0%	0%	0%	1.576	0.592
MAL	Maleku[14]	35	91%	9%	0%	0%	0%	2.992	0.275
MAP	Mapuche [7]	39	15%	41%	21%	23%	0%	0.151	0.916
Ch'orti'	**Maya- Ch'orti'[1]**	**57**	**70%**	**0%**	**25%**	**0%**	**5%**	**2.022**	**0.943**
MKI	Maya- K'iche[4, 5]	27	52%	22%	15%	7%	4%	0.642	0.931
Poqomchi'	**Maya- Poqomchi'[1]**	**65**	**82%**	**6%**	**12%**	**0%**	**0%**	**2.189**	**0.947**
MY	Maya- Yucatec[21]	52	62%	17%	15%	6%	0%	0.885	0.922
MIX	Mixtec[21]	19	79%	11%	5%	5%	0%	1.962	0.825
MOV	Movima [3]	12	0%	8%	75%	17%	0%	2.041	0.894
MOX	Moxo [3]	26	23%	23%	50%	4%	0%	0.514	0.975
NAH	Nahua[21]	84	68%	23%	7%	2%	0%	1.247	0.929
NGO	Ngöbe[10]	46	67%	33%	0%	0%	0%	1.384	0.763
NIC	Nicaragua[18]	163	74%	14%	0%	1%	11%	3.500	0.943
OTO	Otomi[21]	68	40%	25%	29%	6%	0%	0.214	0.967
YUR	Yurimaguas, Peru[8]	52	21%	33%	35%	12%	0%	0.086	0.989

PUR	Purepecha[21]	34	59%	9%	24%	9%	0%	0.867	0.973
QAN	Quechua (Ancash) [13]	33	9%	52%	18%	21%	0%	0.350	0.981
QAR	Quechua (Arequipa)[6]	22	9%	68%	14%	9%	0%	0.774	0.965
QTA	Quechua (Tayacaja)[6]	61	21%	33%	13%	30%	3%	0.359	0.968
RAM	Rama[14]	75	8%	92%	0%	0%	0%	2.187	0.591
SAM	San Martin[6]	22	9%	55%	5%	27%	5%	0.915	0.939
SHA	Shamatari [24]	151	0%	58%	32%	10%	0%	0.668	0.639
TAI	Tainos[11]	24	0%	0%	75%	25%	0%	2.340	0.918
TAR	Tarahumara[21]	15	13%	7%	73%	7%	0%	1.772	0.771
TRI	Triqui[21]	107	72%	28%	0%	0%	0%	1.604	0.548
WAY	Wayuú[16]	30	27%	27%	47%	0%	0%	0.473	0.825
WOU	Wounan[9]	31	29%	19%	48%	3%	0%	0.523	0.912
XAV	Xavante[23]	24	17%	83%	0%	0%	0%	1.646	0.685
YAN	Yanamamõ[17]	129	2%	7%	50%	34%	7%	1.920	0.905
YUC	Yuracare[3]	15	27%	40%	33%	0%	0%	0.242	0.952
ZOR	Zoro [23]	29	21%	3%	14%	62%	0%	2.248	0.759
	Average		26%	34%	24%	15%	1%		

This Study[1], (Batista et al. 1995)[2], (Bert et al. 2004)[3], (Boles et al. 1995)[4], (Torroni et al. 1993a)[5], (Fuselli et al. 2003)[6], (Ginther et al. 1993)[7], (Justice et al. In Press)[8], (Kolman and Bermingham 1997)[9], (Kolman et al. 1995)[10], (Lalueza-Fox et al. 2001)[11], (Lalueza-Fox et al. 2003)[12], (Lewis et al. 2005)[13], (Melton 2008)[14], (Santos et al. 1994)[15], (Melton et al. 2007)[16], (Merriwether et al. 2000)[17], (Nunez et al. 2010)[18], (Rickards et al. 1999)[19], (Salas et al. 2009)[20], (Sandoval et al. 2009)[21], (Schmitt et al. 2004)[22], (Ward et al. 1996)[23], (Williams et al. 2002)[24].

4.1.2. HVS1 Sequencing

The results of the mtDNA HVS1 sequencing for both the Ch'orti' Maya and Poqomchi' Maya are presented in Table 4.2. For the Poqomchi', there are 34 haplotypes characterized by 42 variable sites. For the Ch'orti Maya, there are 30 haplotypes, also with 42 variable sites. These sequences reveal that the majority of A haplogroups in both populations are A2 subhaplogroups, haplogroup C is solely subhaplogroup C1, and haplogroup B is B4, all of which are common in the Americas. One haplogroup within the Ch'orti' population is designated as

other, as it is only distinguished from the Cambridge Reference Sequence (CRS) by one mutational difference. This difference alone is not sufficient to accurately haplotype it, but it is most likely of European origin, since the presence of Asian and African haplogroups would be detected. While haplogroup A2 is the most common haplogroup among the Ch'orti, the most frequent haplotype is C1_10 (16111T, 16223T, 16244A, 16274A, 16298C, 16325C, 16327T) shared by nine individuals. This haplotype is not shared with any other populations reported in the comparative data. These haplotypes are defined with caution, as many diagnostic SNPs lie outside of the HVS1 region needed to confidently identify specific mtDNA haplotypes.

The most common haplotype among the Poqomchi' Maya is designated in Table 4.2 as A2_5 (16111T, 16223T, 16290T, 16319A, 16362C) shared by 13 individuals. This haplotype represents the root HVS1 motif for subhaplogroup A2 with the addition of a common mutation at np 16223. This haplotype is shared with populations from Central America, less frequently in Mesoamerica and the Caribbean, and almost absent in South America. Haplogroup A represents the majority of the haplotype variation for both populations encompassing 70% of the total haplotypes for the Ch'orti and 70.5% for the Poqomchi'.

Table 4. 2. Haplotypes and associated mutational differences compared to the Cambridge reference sequence for the Ch'orti' and Poqomchi' Maya.

			16051	16054	16059	16062	16064	16065	16069	16092	16094	16111	16126	16129	16136	16146	16153	16156	16175	16183	16187	16188	16189	16194	16209	16213	16223	16239	16240	16255	16256	16263	16270	16290	16295	16299	16300	16311	16319	16325	16327	16335	16355	16356	16362	16368	16402	16409
f Ch'orti'	*f Poqomchi'*	Haplotype	A	A	A	A	T	G	C	T	T	C	T	G	T	A	G	G	A	A	C	C	T	A	T	G	C	C	A	A	C	T	C	C	C	C	A	A	T	G	C	T	C	C	T	G	T
1	0	A_1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	T	-	C	-	-	
1	0	A_2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	C	-	-	C	-	-	
1	1	A_3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	G	-	A	-	-	-	-	C	-	-	
0	5	A_4	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	-	-	-	
2	0	A_5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	G	-	T	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
2	0	A2_1	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	A	T	-	-	-	-	C	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	0	A2_2	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	C	
7	4	A2_3	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	0	A2_4	-	-	-	-	-	-	-	-	-	T	-	A	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	T	C	-		
4	13	A2_5	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
7	2	A2_6	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	C	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	1	A2_7	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	C	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
2	0	A2_8	-	-	-	-	-	-	-	-	-	T	C	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	0	A2_9	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	G	-	A	-	-	-	-	C	-	-	
1	0	A2_10	-	-	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	-	-	-	-	T	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	0	A2_11	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	A	T	-	-	-	-	C	-	-	T	-	-	-	C	A	-	-	-	-	C	-	-	
1	0	A2_12	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	A	-	C	-	-	C	C	-	-	T	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	0	A2_13	-	-	-	-	-	-	-	C	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	T	-	-	G	-	A	T	-	-	-	C	-	-	
0	1	A2_14	-	-	-	-	-	T	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	T	-	T	-	-	-	A	-	-	-	-	C	-	-	
0	2	A2_15	-	-	-	-	-	-	-	-	-	T	-	-	-	T	-	-	-	T	-	-	-	-	-	-	T	-	-	-	-	-	-	T	T	-	-	-	A	-	-	-	-	C	-	-	
0	1	A2_16	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	T	-	-	-	A	-	-	-	-	C	-	-	
0	2	A2_17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	T	-	-	-	C	A	-	-	-	-	C	-	-	
0	3	A2_18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	G	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
0	1	A2_19	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	C	-	T	-	-	-	-	-	-	-	T	-	-	-	C	A	-	-	-	-	C	-	-	
0	1	A2_20	-	G	-	-	-	-	-	-	-	T	-	-	-	-	-	G	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	G	-	A	-	-	-	-	C	-	-	
0	1	A2_21	-	-	-	-	-	-	-	-	-	T	-	-	C	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	C	A	-	-	-	-	C	-	-	
0	1	A2_22	G	-	-	-	-	-	C	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
0	2	A2_23	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	G	-	-	A	-	-	-	-	C	-	-	
2	0	A2_24	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	A	-	C	-	-	C	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
0	1	A2_25	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	G	G	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
0	1	A2_26	-	-	G	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	T	-	
0	1	A2_27	-	T	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
0	1	A2_28	-	-	-	-	A	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	G	-	-	A	-	-	-	-	C	-	-	
1	0	A2_29	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	T	-	C	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	0	A2_30	-	-	-	-	-	A	-	-	-	T	-	-	-	-	-	-	-	C	-	-	C	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
1	0	A2_31	-	-	-	-	-	A	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	G	-	A	-	-	-	-	C	-	-	
0	1	A2_32	G	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	T	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
0	1	A2_33	-	-	-	G	-	-	-	-	-	T	-	-	-	-	-	-	G	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	G	-	A	-	-	-	-	C	-	-	
0	2	A2_34	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	T	-	T	-	-	-	-	A	-	-	-	-	C	-	-	
0	4	A2_35	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	G	-	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	G	-	A	-	-	-	-	C	-	-	

			16060	16067	16093	16108	16111	16147	16148	16168	16182	16183	16188	16189	16193	16213	16217	16223	16239	16244	16244	16256	16257	16258	16263	16266	16270	16274	16278	16299	16304	16309	16311	16325	16327	16355	16357	16362	16368	16382	16390	16393
f Ch'orti'	*f Poqomchi'*	Haplotype	G	C	T	C	C	C	C	C	A	A	C	T	C	G	T	C	C	G	G	C	C	A	T	C	C	G	C	C	T	T	A	T	T	C	C	T	T	T	C	G	C
0	1	B4_1	-	-	-	-	-	-	-	-	-	C	-	C	-	-	C	-	-	-	-	-	-	C	C	-	T	-	-		-	-	-	-	-	-	T	C	-	G	A	-	T
0	2	B4_2	-	-	-	-	-	-	T	-	C	C	-	C	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0	1	B4_3	-	-	-	-	-	-	-	-	-	C	-	C	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-
0	1	C1_1	-	-	-	-	-	-	-	-	-	C	-	C	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
0	1	C1_2	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
0	1	C1_3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
0	1	C1_4	-	A	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	C	T	-	-	C	-	-	-	-
0	1	C1_5	-	-	-	-	-	-	-	-	-	C	-	C	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	T	-	-	C	-	?	?	?
0	1	C1_6	C	-	-	-	-	-	-	-	-	C	-	C	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	C	T	-	-	?	?	?	?	?
1	0	C1_7	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	-	T	-	-	-	-	-	-	-	-	C	-	-	C	C	T	-	-	-	-	-	-	-
1	0	C1_8	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	T	-	-	A	-	-	-	-	A	-	-	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
1	0	C1_9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
9	0	C1_10	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	T	-	A	-	-	-	-	-	-	A	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
1	0	C1_11	-	-	-	A	-	-	-	-	-	-	-	C	-	-	-	T	-	-	-	T	-	-	-	-	-	-	-	C	-	-	C	C	T	-	-	-	-	-	-	-
1	0	C1_12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	A	-	-	T	-	-	-	A	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
0	2	C1_13	-	-	C	-	-	-	-	-	-	-	-	C	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	C	T	-	-	-	-	-	-	-
1	0	L2_1	-	-	-	-	-	-	-	T	-	-	T	-	-	-	-	T	-	-	-	-	-	-	-	-	-	T	T	-	-	G	-	-	-	-	-	-	-	-	A	-
1	0	L2_2	-	-	-	-	-	-	-	-	-	-	-	-	T	A	-	T	T	-	-	-	-	-	-	-	-	T	T	-	-	G	-	-	-	-	-	-	-	-	A	-
1	0	Other	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-

Table 4. 3. This table contains the mtDNA sequence diversity and neutrality test statistics for all of the comparative populations in this study.

Population	N	*H*#	k	% Unique	h	π	Tajima's D	Fu's F_s	Source
Ach**é**	63	3	7	0.048	0.204	0.004	-0.399	3.086	Schmitt et. al. 2004
Arsario	28	4	10	0.143	0.725	0.012	1.976	5.745	Melton et. al. 2007
Cayapa	30	8	18	0.267	0.837	0.018	1.155	2.873	Rickards et. al. 1999
Chorotega	24	6	14	0.250	0.670	0.009	-0.585	1.426	Melton 2008
Ciboney	15	10	12	0.667	0.943	0.010	-0.377	**-3.684**	Lalueza-Fox et al. 2003
El Salvador Mixed	90	50	53	0.556	0.919	0.011	**-2.143**	**-26.158**	Salas et al. 2009
Ember**á**	44	20	23	0.455	0.942	0.018	0.459	-4.379	Kolman and Bermingham 1997
Gavi**ão**	28	7	16	0.250	0.862	0.013	0.084	1.969	Ward et. al. 1996
Guayami	39	7	12	0.179	0.819	0.011	1.017	2.337	Melton 2008
Huetar	52	12	19	0.231	0.787	0.013	0.068	-0.030	Melton 2008, Santos et al 1994
Ijka	31	3	12	0.097	0.185	0.005	**-1.584**	-2.961	Melton et. al. 2007
Kogi	21	3	10	0.143	0.524	0.010	0.581	**5.39794***	Melton et. al. 2007
Kuna	63	7	10	0.111	0.592	0.010	1.519	2.775	Batista et. al. 1995
Maleku	14	3	9	0.214	0.275	0.004	**-1.930**	1.633	Melton 2008
Mapuche	39	14	21	0.359	0.916	0.016	0.316	-0.426	Ginther et. al. 1995
Maya- Ch'orti'	56	29	42	0.518	0.943	0.020	-0.980	**-11.251**	This Study
Maya- K'iche	34	18	27	0.529	0.931	0.017	-0.576	-4.901	Boles et al. 1995, Torroni et al. 1993
Maya-Poqomchi'	65	34	42	0.523	0.947	0.015	**-1.494**	**-21.643**	This Study
Maya-Yucatec	52	20	27	0.385	0.922	0.018	-0.114	-3.685	Sandoval et al. 2009
Mixtec	19	10	10	0.526	0.825	0.011	-1.226	-2.130	Sandoval et al. 2009
Movima	12	8	12	0.667	0.894	0.009	-1.093	-2.632	Bert et al. 2004
Moxo	26	20	32	0.769	0.975	0.020	-0.767	**-9.058**	Bert et al. 2004
Nahua	81	41	51	0.506	0.929	0.019	**-1.473**	**-24.974**	Sandoval et al. 2009
Ngobe	46	7	12	0.152	0.763	0.013	1.684	3.388	Kolman et al. 1995
Nicaragua	163	63	68	0.387	0.943	0.017	**-1.631**	**-25.066**	Nuñez et al. 2010
Otomi	68	32	38	0.471	0.967	0.021	-0.436	**-11.579**	Sandoval et al. 2009
Peru (Yurimaguas)	52	41	46	0.788	0.989	0.022	-1.121	**-25.089**	Justice et al. (In Press)
Purepecha	34	23	37	0.676	0.973	0.020	-0.979	**-9.754**	Sandoval et al. 2009
Quechua (Ancash)	33	27	40	0.818	0.981	0.018	-1.482	**-19.791**	Lewis et al. 2004
Quechua (Arequipa)	22	17	25	0.773	0.965	0.015	-1.031	**-8.840**	Fuselli et al. 2003

Quechua (Tayacaja)	61	42	48	0.689	0.968	0.018	-1.389	**-25.268**	Fuselli et al. 2003
Rama	30	7	11	0.233	0.591	0.007	-0.492	0.063	Melton 2008
San Martin (Mixed)	22	15	22	0.682	0.939	0.015	-0.573	**-5.444**	Fuselli et al. 2003
Shamatari	151	6	14	0.040	0.639	0.011	1.203	6.894	Williams et. al. 2002
Tainos	19	11	13	0.579	0.918	0.009	-0.740	**-4.210**	Lalueza-Fox et al. 2001
Tarahumara	15	7	22	0.467	0.771	0.014	-1.210	0.612	Sandoval et al. 2009
Triqui	107	15	27	0.140	0.548	0.013	-0.371	0.182	Sandoval et al. 2009
Wayuú	30	6	17	0.200	0.825	0.016	0.967	4.634	Melton et. al. 2007
Wounan	31	14	29	0.452	0.912	0.020	-0.273	-1.013	Kolman and Bermingham 1997
Xavante	24	4	10	0.167	0.685	0.009	0.514	3.753	Ward et. al. 1996
Yanomam**õ**	129	30	31	0.233	0.905	0.014	-0.476	-8.631	Merriwether et. al. 2000
Yuracare	15	11	22	0.733	0.952	0.020	-0.049	-2.224	Bert et al. 2004
Zoro	29	8	16	0.276	0.759	0.011	-0.203	0.850	Ward et. al. 1996

4.1.3 With-in Group Variation

For comparative purposes the original length of HVS1 was trimmed to include np 16050 to 16383. Measures of diversity for mtDNA sequence data for study population and comparative populations are presented in Table 4.3. The Ijka (0.1849) have the smallest haplotype diversity, while the Aché (0.0038) have the smallest nucleotide diversity. The admixed population from Yurimagas, Peru (0.9887), has the highest gene diversity and nucleotide diversity (0.022). Both the Ch'orti (0.9429) and Poqomchi' (0.9471) have relatively large values for haplotype diversity with the Poqomchi's possessing slightly greater diversity. The Poqomchi' (0.01484) exhibit a moderate nucleotide diversity measure, but the Ch'orti (0.01993) have much higher nucleotide diversity. Both of the other Maya populations exhibit

similar diversity measures with only slightly smaller haplotype diversity and intermediate values of nucleotide diversity.

4.1.4. Among-population Comparisons

4.1.4.1. AMOVA

Three hierarchical models were tested using AMOVA, one grouped according to major Maya region, including the four Maya populations, one grouped according to major geographic region, and the last according to major language group. Only loci missing less than 5% of the data were included, leaving a total of 334 bases used in these analyses (np 16050-16383). The results of the first AMOVA, grouping the Maya populations into major geographical region are given in Table 4.4. The analysis reveals little variation among the populations within-groups and among-groups, with 94.55% of the variation resting within the populations. This is further illustrated by the low fixation indices, ϕ_{ST} (0.05450) (among populations among groups), ϕ_{SC} (0.06476) (among populations within groups), and ϕ_{CT} (-0.01097) (populations among groups). As mentioned, negative values are possible when compared to a null model.

Table 4. 4. Results of the AMOVA analysis in which populations were grouped by major Maya region. Statistically significant results are noted with an asterisk.

Source of Variation	D of f	SSD	σ^2	% of σ^2
Among groups	2	24.366	-0.03549	-1.1
Among Populations within groups	1	12.518	0.21185*	6.55
Within Populations	203	621.093	3.05957*	94.55
Total	206	657.977	3.23594	

The results of the second AMOVA are presented in Table 4.5, where populations were grouped by major geographic area. There are moderate levels of population structure across geographic regions, the majority of the variation still lies within populations, but more variation is explained by covariation of haplotypes among populations within groups and among the groups. Additionally, the fixation indices reveal the same pattern, with moderate ϕ_{ST} (0.28221), ϕ_{SC} (0.18387), and ϕ_{CT} 0.12049), with the ϕ_{CT} explaining the degree of fixation due to the covariance of haplotypes among groups. In each case the fixation indices and hierarchical variance scores are statistically significant ($P<0.0001$) based on the probability of observing the same or lower measure for each statistic.

Table 4. 5. This table displays the results of the AMOVA analysis in which populations were grouped by major geographical region. Statistically significant results are noted with an asterisk.

Source of Variation	D of f	SSD	σ^2	% of σ^2
Among groups	3	699.993	0.44478*	12.05
Among Populations within groups	39	1157.47	0.59697*	16.17
Within Populations	1964	5203.91	2.64965*	71.78
Total	2006	7061.37	3.69139	

The results of the third AMOVA are presented in Table 4.6, where populations were grouped by language family. Like in the AMOVA of geography, the majority of the variation lies within populations rather than among major language families, but with moderate to large levels of population structure due to covariance among populations and groups. Additionally, the fixation indices reveal the same pattern, with high ϕ_{ST} (0.26868) and ϕ_{SC} (0.18069) and low ϕ_{CT} (0.10739). In each case the fixation indices and hierarchical variance scores are statistically significant (P<0.0001) based on the probability of observing the same or lower measure for each statistic.

Table 4. 6. This table displays the results of the AMOVA analysis in which populations were grouped by major language area. Statistically significant results are noted with an asterisk.

Source of Variation	D of f	SSD	σ^2	% of σ^2
Among groups	17	1005.64	0.37967*	10.74
Among Populations within groups	21	578.527	0.5702*	16.13
Within Populations	1641	4242.66	2.58541*	73.13
Total	1679	5826.83	3.53528	

4.1.4.2 MJ Network Analysis

The resulting network generated through Median Joining network analysis is shown in Figure 4.1. This figure includes all four major Native American mtDNA haplogroups (A, B, C, D). Haplogroup X is absent from all four Maya populations, and therefore, was not included in the

network. Haplogroup A2 is the most common haplogroup among all four populations, which is evidenced in the large founder node with the haplogroup A cluster. It is important to note that this haplotype is not found among the K'iche Maya, but is the most frequent in the Poqomchi', Ch'orti, and Yucatec Maya. There is a clear star-like cluster surrounding the founder node for haplogroup A indicative of a population undergoing expansion. However, there are several satellite nodes and long branches with smaller star structures indicating the presence of deep lineages. The network analysis highlights some major differences between the Maya populations. First, the K'iche occupy their own satellite node for cluster D, the Ch'orti have a satellite node on haplogroup C, and the Poqomchi' have an excess of singleton mutations on haplogroup A. Since the majority of haplogroups are in A, a separate network analysis was performed for Haplogroup A for just the Poqomchi' and then for the Ch'orti. These are shown in Figure 4.2a and 4.2b below.

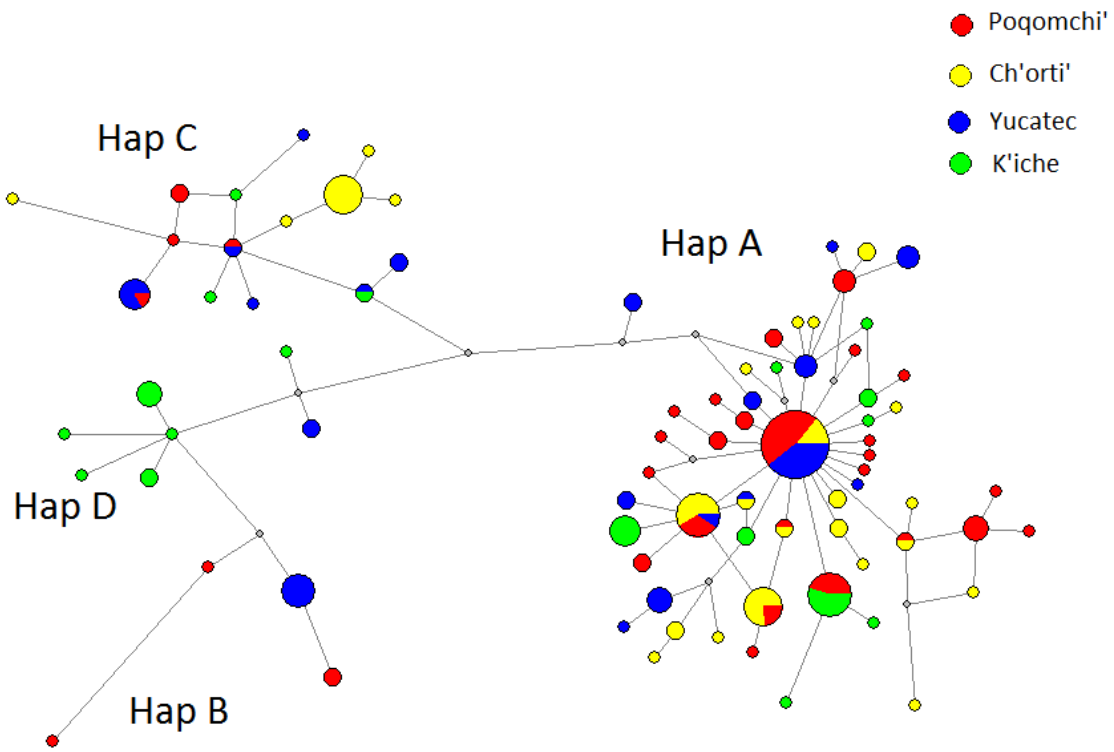

Figure 4. 1. Median-Joining Network analysis for the Maya populations in the current study and comparative populations. Only the four major Native American haplogroups were included in the calculations.

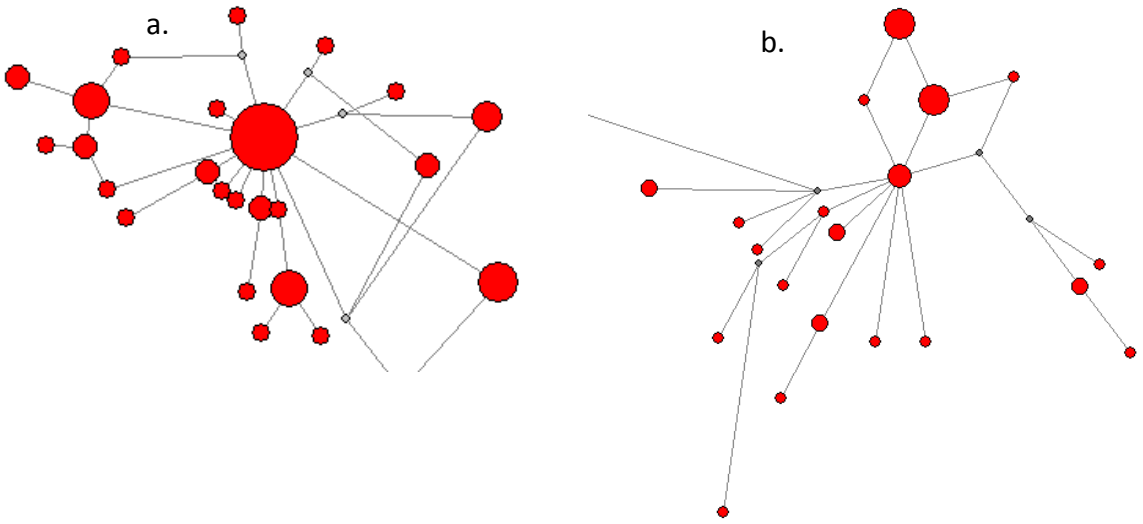

Figure 4. 2. Haplogroup A for the Poqomchi' Maya (a) and the Ch'orti' Maya (b). Branches connecting haplogroup A to the other founder haplogroups are missing nodes.

The networks of Poqomchi' and Ch'orti' haplogroup A highlight additional intra-population structure. The Poqomchi' possess several unresolvable clusters due to reticulations, which break up the general star-like cluster of the total Maya populations. Both populations exhibit a network structure with evidence of past expansion followed by a numerical reduction, as evidenced by the presence of star-like clusters and long fragmented branches. See figure 4.3 for a basic simulated network analysis that includes a deep population expansion, followed by a recent reduction. This figure resembles the Poqomchi' network with the star-like cluster and few fragmented branches.

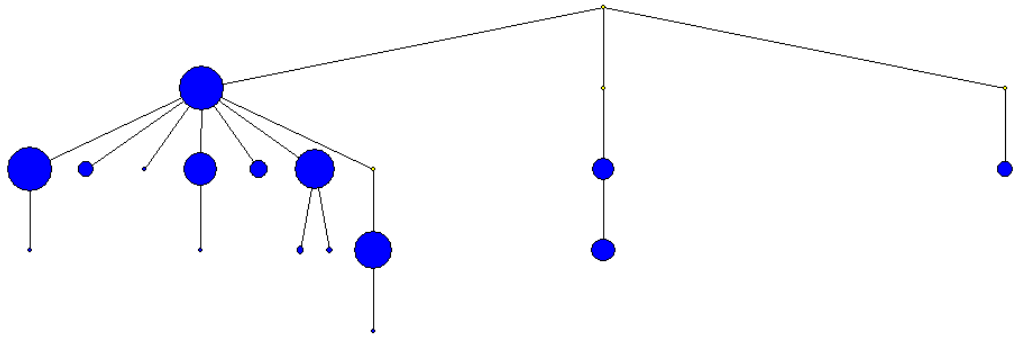

Figure 4. 3. Simulated median-joining network predicted by population expansion followed by a population reduction

4.1.4.3 Multi-dimensional Scaling Plots

Both monotonic and linear multi-dimensional scaling plots were generated for two and three dimensions. A three dimensional monotonic MDS, shown in Figure 4.4, resulted in the lowest stress value (0.05616), indicating a good fit of the projection to the original distance matrix. The Maya populations tend toward the center of the plot while the other Mesoamerican populations are more widely scattered. Also, the Central American populations cluster closely with the Andean populations while the Amazonian populations tend to be more dispersed across the MDS. This pattern mimics patterns previously seen where Andean populations show a close relationship with Central American populations, higher variation within populations in the Andes and greater variation among groups in the Amazon. This has been cited as evidence that South America was peopled from Central America along the Western coast and then east into Amazonia (Lewis et al. 2007). These populations, clustered on the right side of the three dimensional plot, also share high relative frequencies of haplogroup

B, while the left side tends to have higher frequencies of haplogroup A. Also, the Caribbean populations, Tainos (TAI) and Ciboney (CIB), cluster closely due to their high frequencies of haplogroup C and D and lack of haplogroup B.

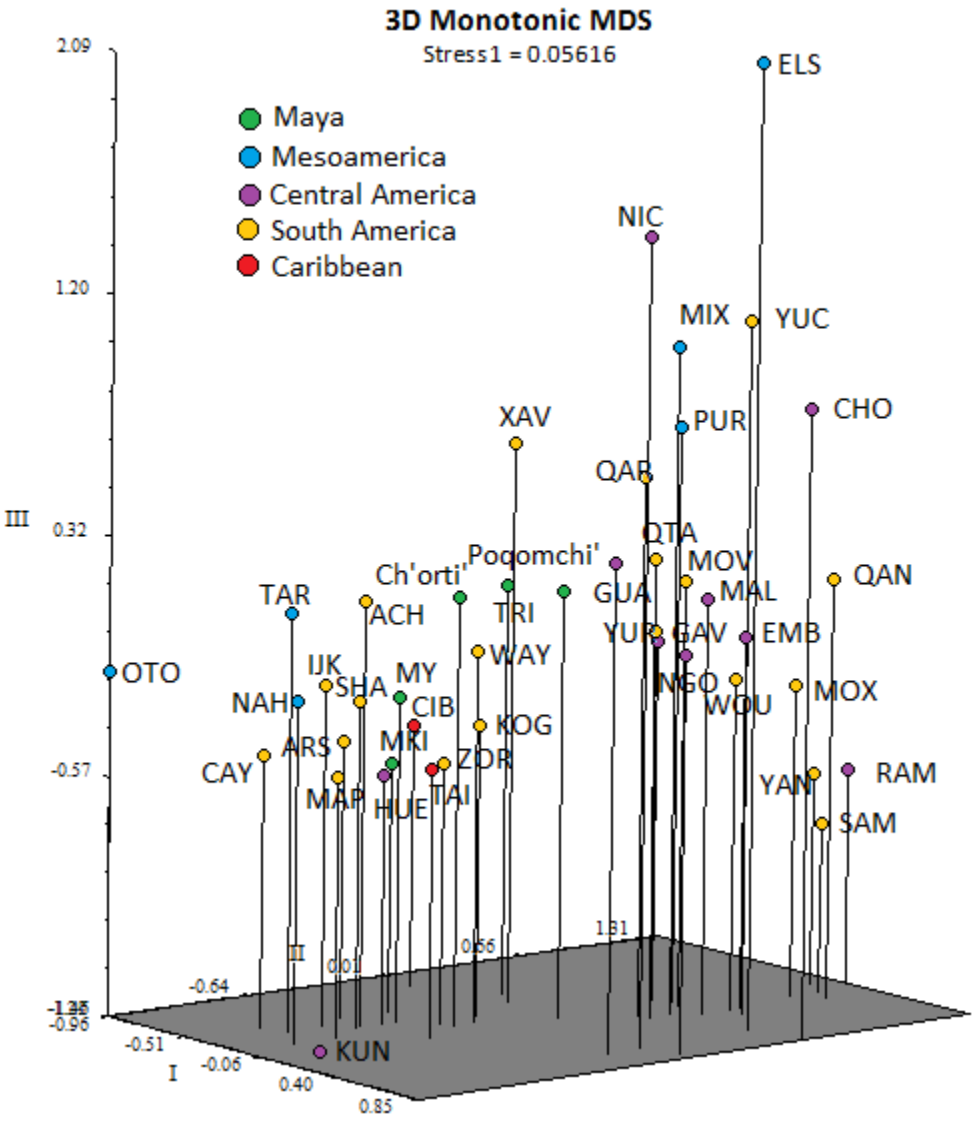

Figure 4. 4. Three dimensional MDS plot with the lowest stress value for the mtDNA kimura-2p genetic distance matrix ($p<0.05$).

4.1.4.4. Neighbor- Joining Tree

A neighbor-joining tree (NJT) constructed using the Kimura 2-p distances is shown in Figure 4.5. Overall, population clusters based on genetic distance do not conform to the geographic or linguistic relationship among the populations, but a few patterns are apparent in the plot. First, the Caribbean populations cluster together on a single branch. Also, South American and Central American Chibchan populations with high frequencies of haplogroup B share a common branch. Other populations are scattered throughout the tree. A cophenetic distance matrix was generated from the NJT and a mantel test performed comparing the cophenetic matrix to the original distance matrix. The mantel test revealed a low, but significant correlation between the two distance matrices ($r = 0.51971$, $p < 0.05$). This indicates that the NJT is not an accurate representation of the relationship among the populations.

4.1.5. Forces of Evolution

4.1.5.1. Neutrality test statistics

Neutrality test statistics are used to quantify the probability of a population expansion or genetic drift using sequence data. The results of the neutrality test scores computed for this study, Tajima's D and Fu's F_S are presented in Table 4.2. For both tests, the Ch'orti and Poqomchi' express large negative values, indicative of a population expansion, with both scores being statistically significant for the Poqomchi' (D= -1.49353 and Fs= -21.64273) and the F_S (-

11.2544) is significant for the Ch'orti'. The admixed population from El Salvador expressed the highest negative values for both scores. Because Arlequin only tests for the probability of getting a score less than the observed, there are no statistically significant positive scores for Tajima's D. However, the many of the southern Central American Chibchan populations and the Amazonian populations have the highest scores for Tajima's D. Further analyses are needed to determine if these results are truly indicative of a pattern of genetic drift. Among these populations, the Kogi have a large positive and statistically significant Fu's F_s score, supporting the implications of the D scores and findings of previous investigations on these populations (Melton 2008).

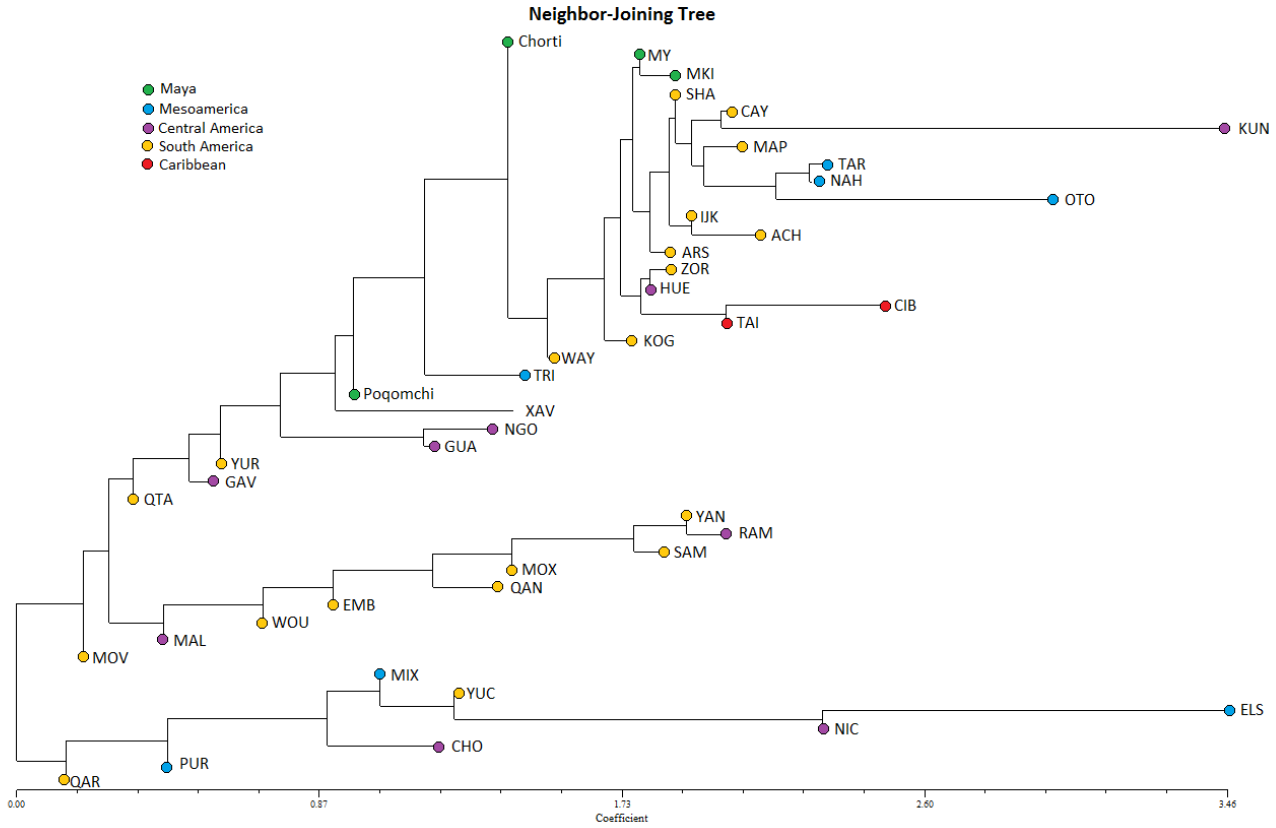

Figure 4. 5. This figure displays the Neighbor-Joining tree created using kimura-2p distance for all of the populations examined in this study.

4.1.5.2 Mismatch Analysis

The results of the pairwise mismatch analysis are given in Figures 4.6, 4.7, and 4.8. For the total Poqomchi', total Ch'orti', and pooled Maya populations, respectively (focus populations and K'iche and Yucatec), the histograms including all haplogroups display a bimodal distribution. This distribution may be misleading, as studies have shown that each haplogroup entered the Americas at the same time, and then evolved, without recombination, separately. Therefore, combining these haplogroups in a single mismatch will inflate the number of pairwise differences relative to the actual number of accrued mutations. Therefore, these bimodal distributions are not indicative of recent bottlenecks as would normally be inferred. Therefore, each figure also displays the histogram of pairwise differences for the two most frequent haplogroups (A and C) separately (haplogroup B is also shown for the Poqomchi'). In all instances, haplogroup A, the most frequent haplogroup within each Maya population, displays a unimodal distribution. The Poqomchi' has the lowest mean pairwise differences (MPD) for this haplogroup (MPD=2.417), while the Ch'orti' have a MPD of 3.163, and all Maya have a MPD of 2.867. Haplogroup B is displayed for the Poqomchi' Maya, but as only four individuals possess haplogroup B, little inference can be made and the resulting bimodal plot is due to small sample size. Haplogroup C displays a multimodal distribution for each population individually and pooled Maya, with both the Ch'orti and pooled historgrams displaying a peak at zero differences. This is indicative that, while haplogroup C is only found in moderate frequencies, it exhibits currently stable population growth.

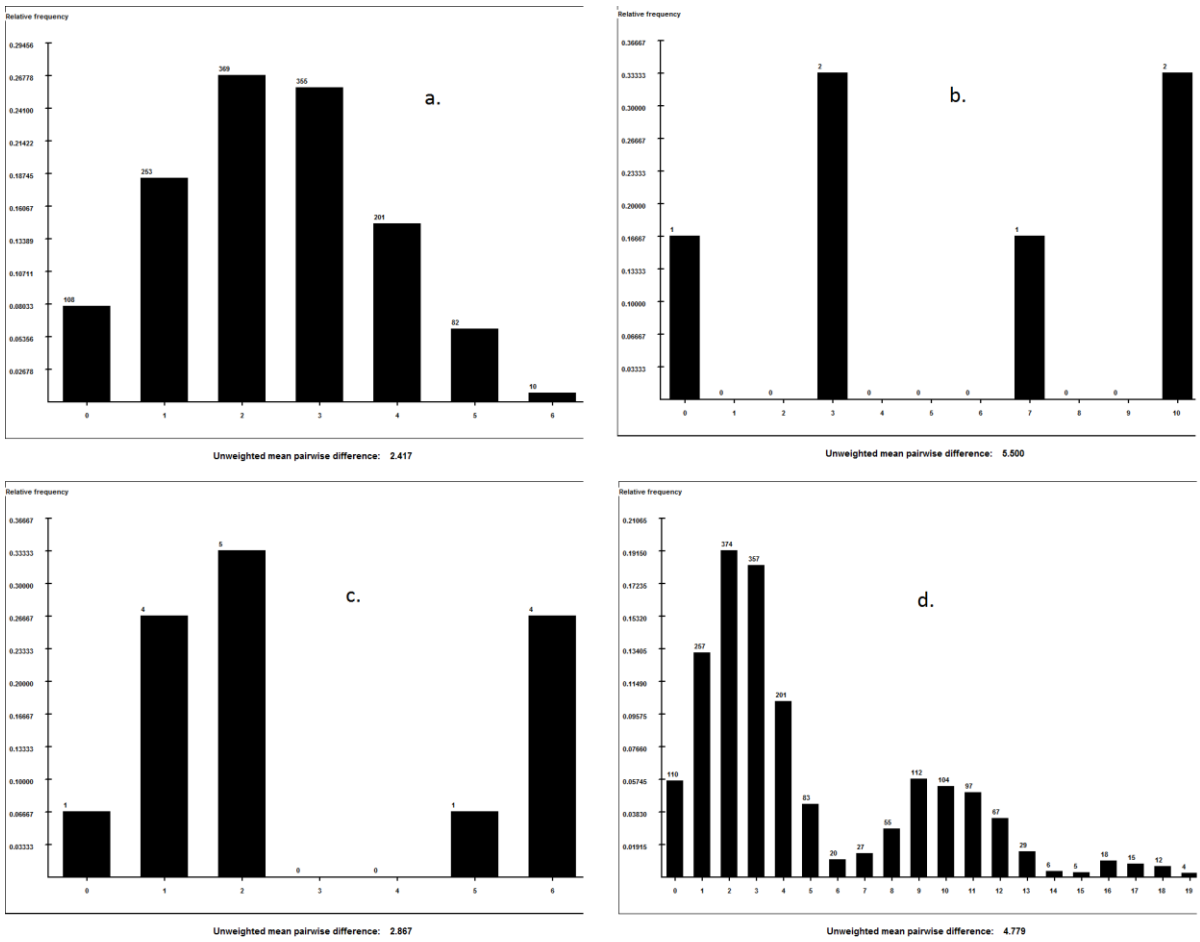

Figure 4. 6. Each histogram represents the frequency and proportionate frequency of pairwise differences among individuals in the Poqomchi' Maya sample for haplogroups: a) Hapgroup A, b) Haplogroup B, c) Haplogroup C, and d) pooled haplogroups.

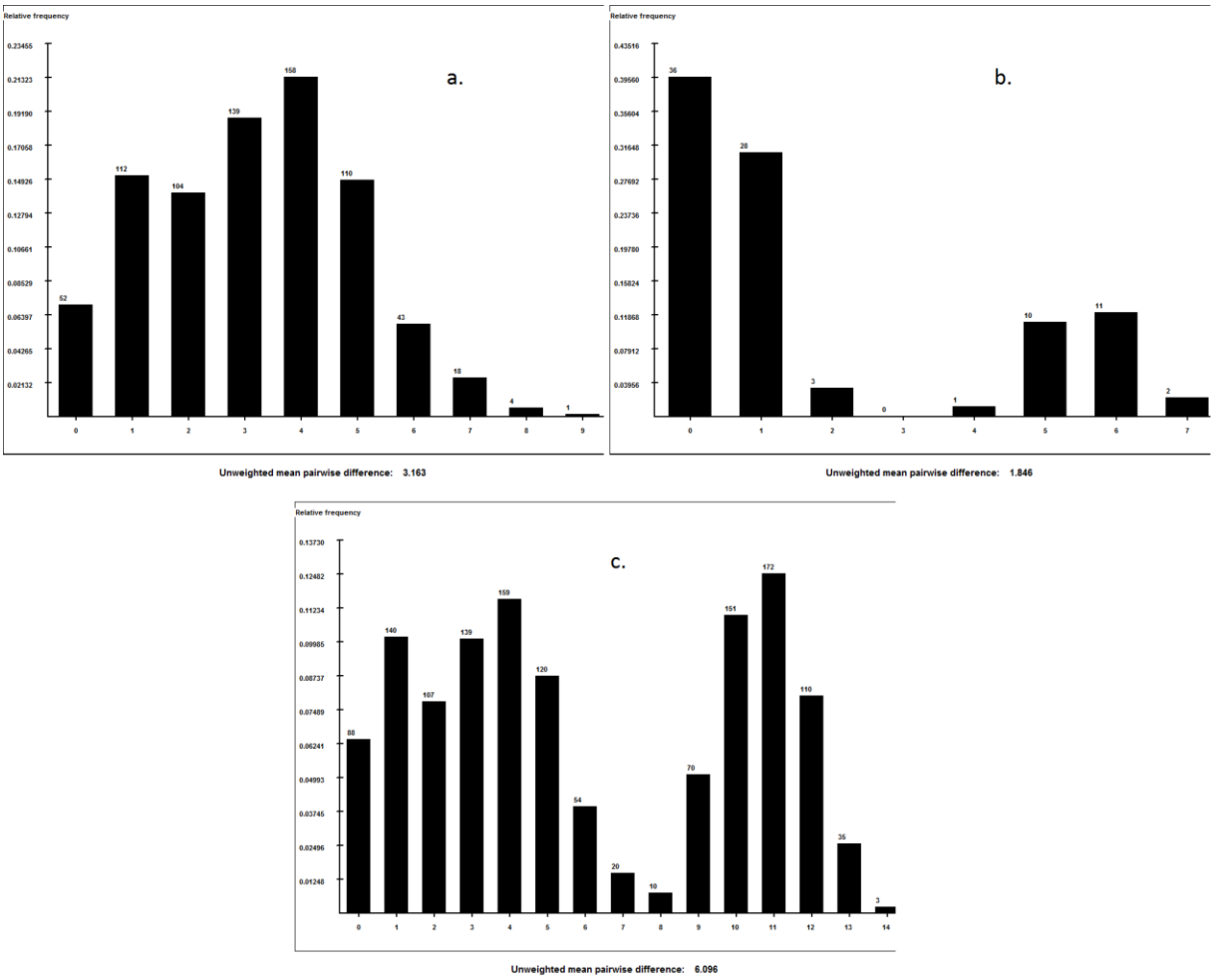

Figure 4. 7. Each histogram represents the frequency and proportionate frequency of pairwise differences among individuals in the Ch'orti' Maya sample for haplogroups: a) Hapgroup A, b) Haplogroup C, and c) pooled haplogroups.

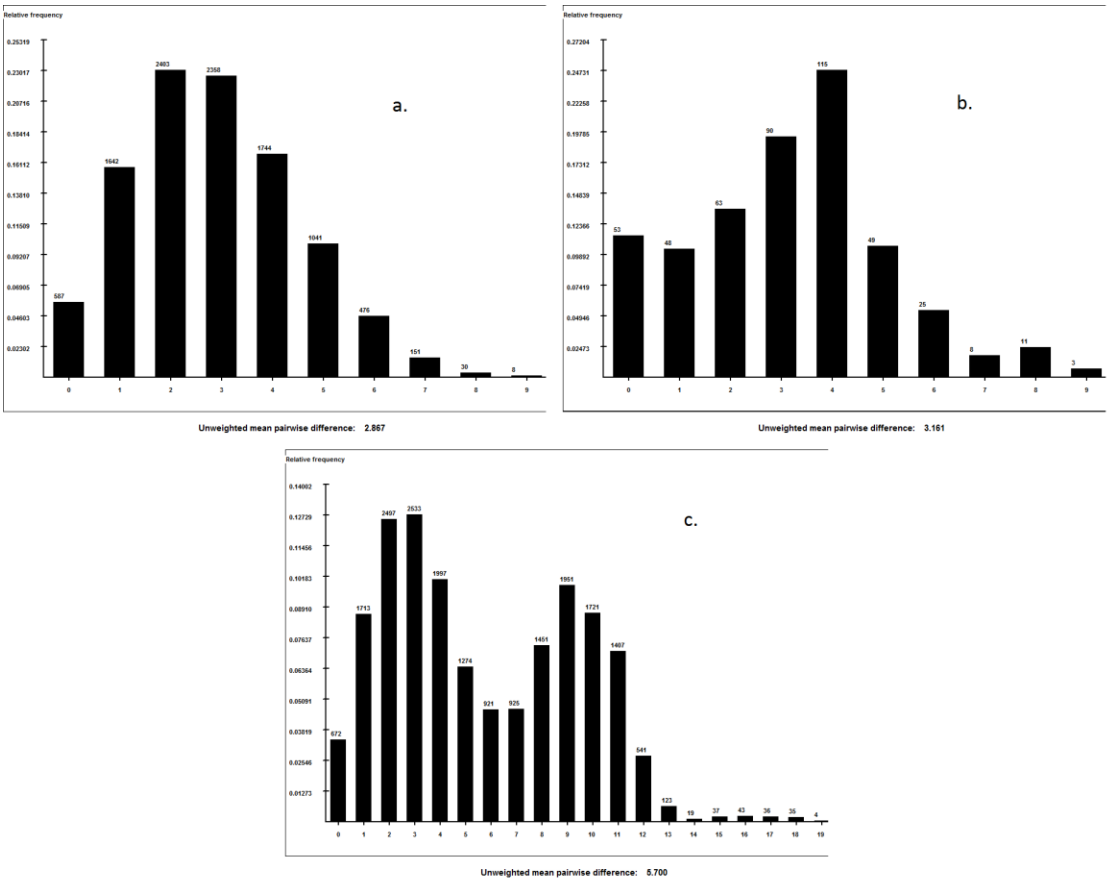

Figure 4. 8. Each histogram represents the frequency and proportionate frequency of pairwise differences among individuals in the pooled Maya (Poqomchi', Ch'orti', K'iche, and Yucatec) samples for haplogroups: a) Hapgroup A, b) Haplogroup C, and c) pooled haplogroups.

4.5.1.3 Diversity vs. r_{ii}

The results of the regression or gene diversity (*h*) on distance from the centroid (r_{ii}) of haplogroup frequency differences is given in Figure 4.9. A weak relationship exists between the two measure of variation (r^2= 0.1782) indicating that the assumption of an isolation by distance model does not hold true for these populations. All of the Maya populations lie above the regression line indicating that they have higher than expected diversity and have experienced

gene flow, with the Poqomchi' and Ch'orti Maya having the highest deviation above this line for the Maya. Both highly admixed populations from El Salvador and Nicaragua lie above the line, as expected. The isolated Chibchan populations, lie below or close to the regression line supporting the findings of other tests which indicate genetic drift within these populations.

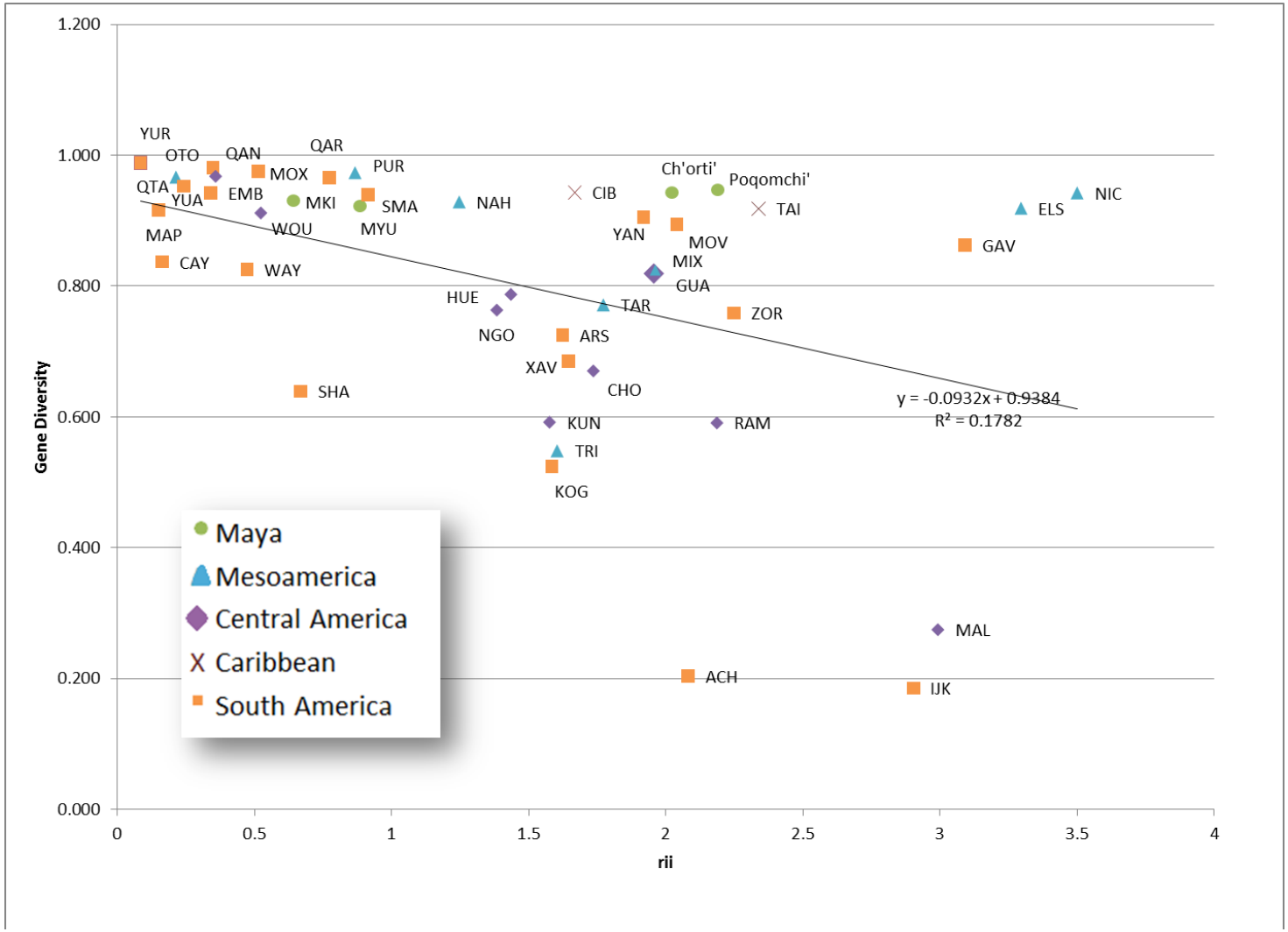

Figure 4. 9. Scatter plot and resulting predicted regression line for the comparison of gene diversity and r_{ii} among the comparative populations.

4.1.6 Phylogeography

4.1.6.1 Mantel Tests

A Mantel Randomization test was conducted comparing geographic distance and genetic distance, which resulted in a non-significant and negligible correlation (r = 0.0317, p = 0.669). The comparison of linguistic distance and genetic distance also resulted in a very small and non-signficant correlation (r = 0.01578, p = 0.6541). There is no clear relationship between geography or language to genetics among these populations.

4.1.6.2 SAMOVA

Table 4. 7. Results of the SAMOVA analysis for each number of specified groups.

Number of Groups	**F_{ct}**
2	0.21032
3	0.22104
4	**0.23047**
5	0.2194

The SAMOVA analysis that predicted membership for four groups provided the highest F_{ct} score, explaining 23.05% of the variation among the groups. The F_{ct} scores obtained from predicting two, three, and five groups are displayed in Table 4.7. For the four group model with the highest score, group one included populations from Meso-, Central and Northern South America (Poqomchi' Maya, Ch'orti' Maya, Yucatec Maya, Mixtec, K'iche Maya, Cayapa, Arsario, Ijka, Kogi, Huetar, Kuna, Ngobe, Triqui, Purepecha, Nahua, Chorotega, Guatuso Maleku,

Guaymi, El Salvador Mixed, Nicaragua Mixed), group two included the two Caribbean populations, Ciboney and Tainos. Group three includes populations from South America and populations from Meso- and Central America that had greater genetic distance from adjacent populations (Shamatari, Wayuu, Yanamamo, Mapuche, Gavaio, Zoro, Emberra, Wounan, Peru, Moxo, Yuracare, Movima, Arequipa Quechua, Tayacaja Quechua, San Martin Mixed, Ancash Quechua, Tarahumara, Otomi, and Rama). The last group included only the Ache and Xavante. While these groups coincide with the highest fixation indices, they do not match those groupings suggested by the MDS plots of genetic distance.

4.1.6.3 Monmonier's Maximum Difference Algorithm

The results of the Monmonier's Maximum Difference Algorithm are displayed in Figure 4.10. The first barrier to gene flow detected by the algorithm is highlighted in red, and first separates the Caribbean populations from the others, as well as identifies the Otomi as an outlier among the Mesoamerican populations and the Kogi among the South Americans. The second barrier is highlighted in yellow, which separates the population from Yurimaguas, Peru from all other populations. The third barrier is green and detects a barrier, highlighting the Rama as separate from other Central Americans. The blue barrier is the fourth barrier detected by the algorithm, and separates the rest of South America from Meso- and Central America. The final barrier, shown in purple, points out the restricted gene flow between Central America and Amazonia.

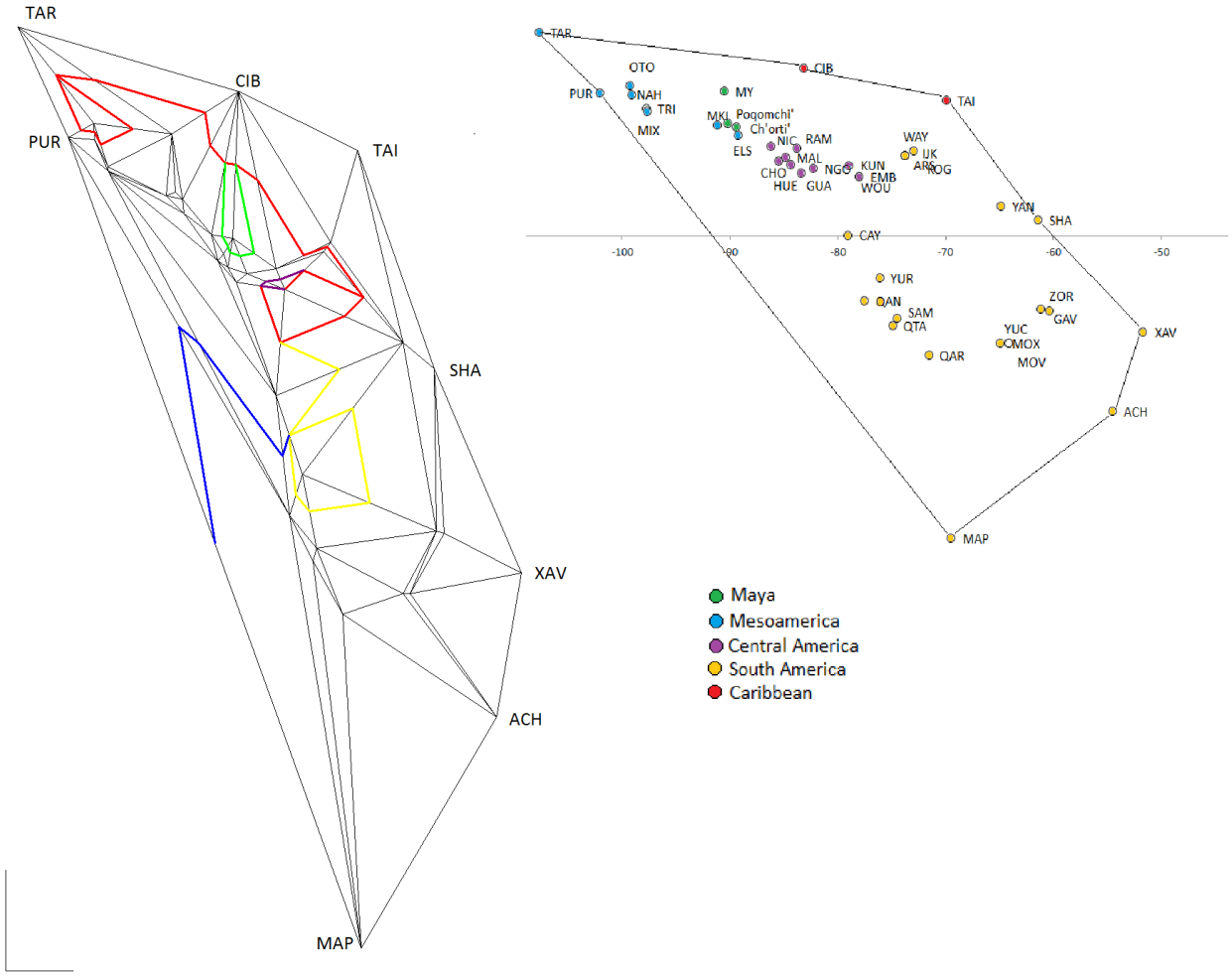

Figure 4. 10. Possible barriers to gene flow predicted by the Monmonier's Maximum Difference Algorithm; 1st red, 2nd yellow, 3rd green, 4th blue, 5th purple.

4.6.6.4 Interpolated Genetic Landscape

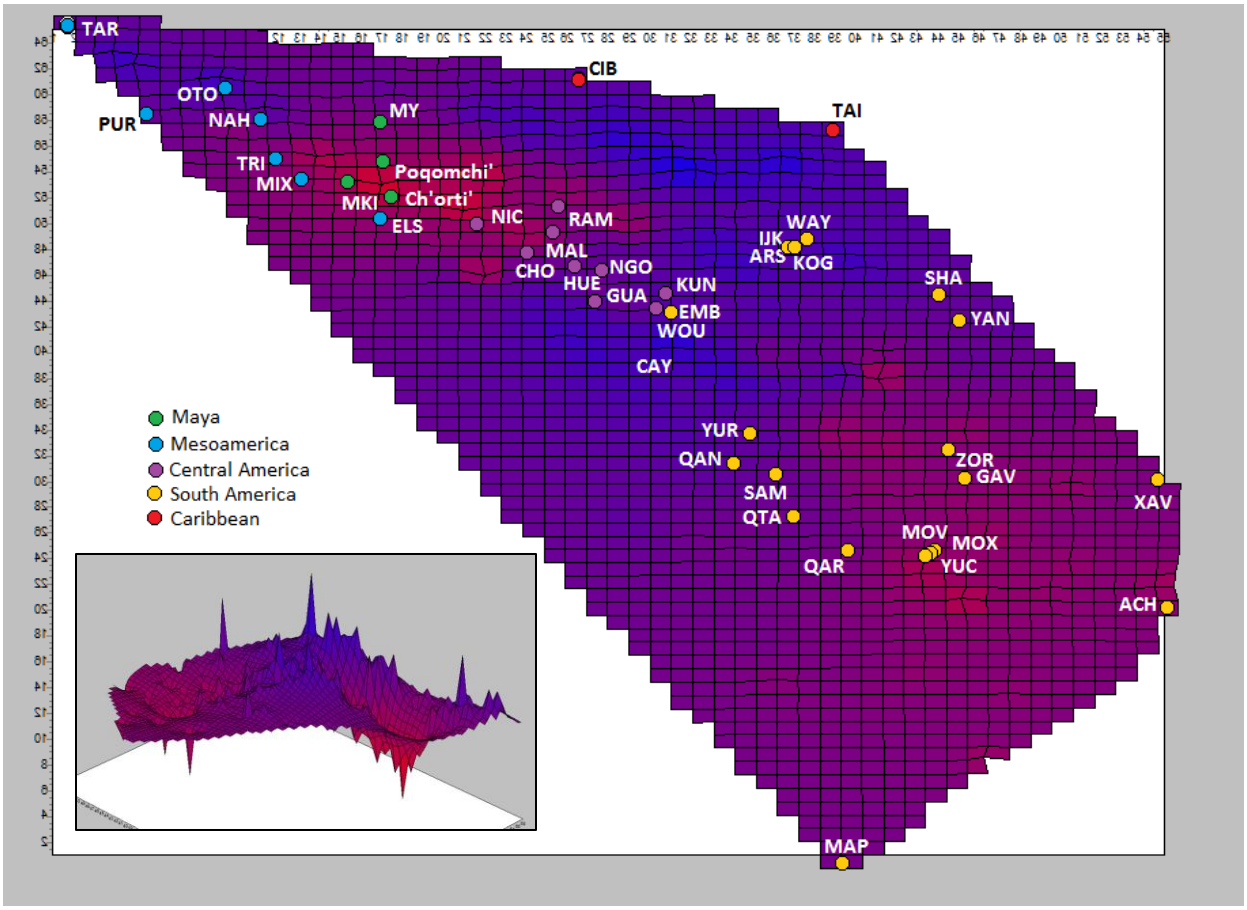

Figure 4. 11. This Interpolated genetic landscape was created using residual scores from the regression of geographic distance on kimura-2p distance using mtDNA sequence data.

The interpolated genetic landscape (GL) provides a plot of the latitude and longitude of the approximate collection sites for the populations, with the z-axis representing the residual scores from the genetic distance *versus* geographical distance between two populations in the area. The blue peaks shown in Figure 4.11 are indicative of genetic drift, as these represent

greater genetic distance than expected. While, the red nadirs highlight instances of gene flow, as they highlight decreased distance between two populations compared to geographic distance. In this instance, there is a clear pattern of greater population similarity, and therefore likely gene flow, among the Mesoamerican populations and again among the western South American populations. However, there is greater diversity among populations in Central America and eastern South America.

4.2. Y CHROMSOME ANALYSES

4.2.1 Y Chromosome Motifs and Haplogroups

Each sample was characterized for nine Y STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439), four Y-SNPs (Q-M242, Q1-P36.2, Q1a3a1-M3, and R1b-M269), and the remaining unknown haplogroups were typed using the STR profiles and Whit Athey's haplogroup predictor. Table 4.8 presents the results of the Y chromosomes STR and SNP analyses for both the Ch'orti' and Poqomchi' Maya. Ignoring missing data, the Ch'orti Maya (N=19) exhibit 18 unique haplotypes with only one haplotype repeated in the population. The Poqomchi' Maya (N=25) exhibit 25 unique haplotypes with no repetitive haplotypes. The most common haplogroup among the Ch'orti and Poqomchi' Maya is haplogroup Q (73.7% and 96% respectively), with the majority of these being the Native American specific Q1a3a1 (73.7% and 84% respectively). In addition to haplogroup Q, the Poqomchi' Maya have one unknown haplogroup that could not be identified using the Y-SNP

probes or Whit Athey's haplogroup predictor. The Ch'orti' Maya exhibit several other haplogroups J (15.8%), I (5.3%), and E (5.3%), which provide evidence of non-Native American admixture.

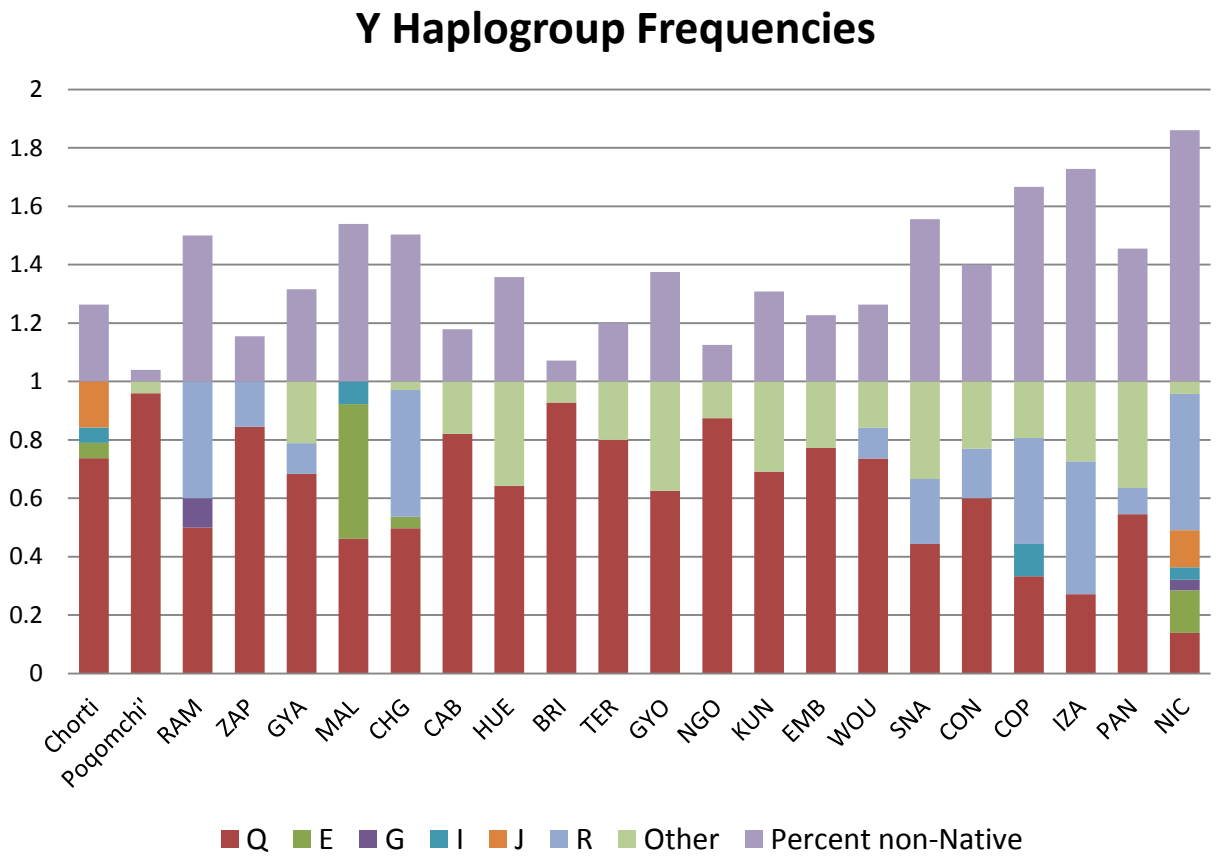

Figure 4. 12. Reported haplogroup frequencies for each comparative population. Those populations for which no frequencies were reported were left out of this figure.

Table 4. 8. Haplotypes and associated frequencies present in both of the focus populations, Ch'orti' and Poqomchi'.

Sample	Haplogroup	Ch'orti'	Poqomchi'	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439
E1	E1b1b	1	0	13	15	33	23	10	11	13	10	12
I1	I2a1	1	0	14	15	31	23	10	11	13	10	12
J1	J1	1	0	14	12	26	23	10	11	12	10	11
J2	J2a1	1	0	14	12	26	25	11	11	12	9	12
J3	J2b	1	0	16	18	29	22	10	11	12	10	12
Q1	Q	0	1	13	17	30	23	10	15	14	11	13
Q2	Q	0	1	13	13	30	24	10	13	13	11	12
Q3	Q	0	1	14	14	30	23	10	14	11	11	13
Q4	Q1a3a	0	1	12	12	24	19	13	7	14	7	10
Q5	Q1a3a	1	0	13	12	24	24	10	13	13	11	13
Q6	Q1a3a	1	0	13	12	24	24	10	14	12	11	11
Q7	Q1a3a	1	0	13	12	26	24	10	13	13	11	13
Q8	Q1a3a	0	1	13	12	30	24	10	13	13	11	12
Q9	Q1a3a	0	1	13	12	30	24	10	13	14	11	12
Q10	Q1a3a	1	0	13	13	24	23	10	11	13	10	12
Q11	Q1a3a	0	1	13	13	30	24	7	14	14	11	12
Q12	Q1a3a	1	0	13	13	30	24	10	13	12	11	12
Q13	Q1a3a	0	1	13	13	30	25	10	13	14	11	11
Q14	Q1a3a	0	1	13	13	31	24	10	13	13	11	13
Q15	Q1a3a	1	0	13	13	?	26	10	14	12	11	11
Q16	Q1a3a	1	0	13	14	30	24	10	13	13	11	11
Q17	Q1a3a	1	0	13	14	30	24	11	13	13	11	13
Q18	Q1a3a	0	1	13	14	31	23	10	14	15	11	13
Q19	Q1a3a	2	0	13	14	31	24	10	13	13	11	12
Q20	Q1a3a	0	1	13	14	31	24	11	14	13	11	12
Q21	Q1a3a	1	0	13	14	32	?	10	14	12	11	11
Q22	Q1a3a	0	1	13	14	?	24	10	13	13	11	13
Q23	Q1a3a	1	0	13	15	26	24	10	13	13	11	11
Q24	Q1a3a	0	1	13	17	?	24	10	13	14	12	11
Q25	Q1a3a	0	1	13	?	?	24	10	14	13	11	11
Q26	Q1a3a	0	1	13	?	?	24	10	14	13	11	13
Q27	Q1a3a	0	1	13	?	?	25	10	14	13	11	10
Q28	Q1a3a	1	0	13	?	?	24	10	13	13	11	12
Q29	Q1a3a	0	1	14	14	30	24	10	13	13	12	14
Q30	Q1a3a	0	1	14	14	30	24	10	13	14	11	12
Q31	Q1a3a	1	0	14	15	32	23	11	13	13	11	11
Q32	Q1a3a	0	1	14	15	33	24	10	13	12	11	12
Q33	Q1a3a	0	1	15	12	31	25	9	14	12	11	12
Q34	Q1a3a	0	1	15	14	30	25	9	13	12	11	12
Q35	Q1a3a	0	1	15	14	31	24	9	14	12	11	12
Q36	Q1a3a	0	1	15	17	31	24	9	14	12	11	12
Q37	Q1a3a	0	1	15	?	?	24	10	14	12	11	12
Q38	unknown	0	1	14	17	?	24	10	10	13	10	11

4.2.2. Within-population Variation

The results of the diversity measures are shown in Table 4.9 and Table 4.10 along with the number of loci used for each measurement. In Table 4.9 the gene diversity is given for each locus studied (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS438). The most variable locus for both populations is DYS389I/II and DYS438 exhibits the lowest gene diversity. The Bribri exhibit the smallest value for average locus diversity (0.244), while Izalco have the highest (0.701). The Ch'orti' and Poqomchi' exhibit moderate locus diversity (0.539 and 0.569 respectively). The Guaymi-Osso (0.679) have the lowest haplotype diversity and both the Izalco and San Alejo exhibit the highest haplotype diversity (1.000 for both). The Ch'orti and Poqomchi' both have relatively high haplotype diversity (0.983 and 0.990 respectively) compared to the total mean (0.915). The measures of mean pairwise difference (π) exhibit the same pattern as the gene diversity with Bribri (1.463) having the lowest, Izalco (6.309) the highest and the Poqomchi (3.980) and Ch'orti' (3.771) having intermediate scores and a total mean of 3.913 across all populations in the study. Figure 4.12 provides a histogram of the various haplogroup frequencies in the populations used in this study. The Poqomchi' exhibit the highest frequency of haplogroup Q among all of the populations, while the Ch'orti' have moderate frequencies of Q. As a result, the Poqomchi' also have the lowest non-native admixture. The Ch'orti' have the highest frequency of haplogroup J, and share haplogroups I and E with only a few other populations in the study. Many populations did not report their haplogroups, or were unable to identify haplogroups other than Q preventing definitive interpretation of differences in non-native gene flow.

Table 4. 9. Gene diversity for the two foci populations at each locus

	Ch'orti'	Poqomchi'	Average
DYS19	0.433	0.630	0.531
DYS389I	0.801	0.795	0.798
DYS389II	0.876	0.608	0.742
DYS390	0.641	0.517	0.579
DYS391	0.281	0.470	0.375
DYS392	0.632	0.630	0.631
DYS393	0.491	0.730	0.611
DYS438	0.485	0.297	0.391
DYS439	0.649	0.707	0.678

Table 4. 10. Y chromosome within-group variation, including number of haplotypes, average locus diversity, haplotype diversity (*H*), and MPD (π).

Populations	Code	N	# of Loci	# haplotypes	Locus Diversity	*H*	π	References
Ch'orti'	Chorti	19	7	17	0.539	0.983	3.771	This Study
Poqomchi'	Poqomchi	25	7	23	0.569	0.990	3.980	This Study
Rama	RAM	20	8	14	0.518	0.968	4.142	Melton 2008
Huetar Zapaton	ZAP	13	8	12	0.524	0.987	4.192	Melton 2008
Abrojo Guaymi	GYA	19	8	14	0.531	0.959	4.246	Melton 2008
Maleku	MAL	13	8	8	0.487	0.897	3.897	Melton 2008
Chortega	CHG	23	8	19	0.616	0.984	4.929	Melton 2008
El Salvador mixed	ESM	150	9	105	0.607	0.986	5.462	Matamoros et al. 2010
Honduras mixed	HDM	128	9	89	0.613	0.988	5.520	Matamoros et al. 2010
Cabecar	CAB	28	6	12	0.412	0.717	2.471	Ruiz-Narváez et al. 2005
Huetar	HUE	28	6	16	0.553	0.942	3.317	Ruiz-Narváez et al. 2005
Bribri	BRI	29	6	8	0.244	0.712	1.463	Ruiz-Narváez et al. 2005
Teribe	TER	15	6	7	0.265	0.724	1.590	Ruiz-Narváez et al. 2005
Guaymi-Oso	GYO	8	6	3	0.512	0.679	3.071	Ruiz-Narváez et al. 2005
Ngobe	NGO	32	7	18	0.399	0.923	2.794	Ascunce et al. 2009
Kuna	KUN	26	6	8	0.489	0.825	2.932	Ascunce et al. 2009
Embera	EMB	22	7	12	0.325	0.909	2.273	Ascunce et al. 2009
Wounan	WOU	19	7	12	0.403	0.942	2.819	Ascunce et al. 2009
San Alejo	SNA	10	6	10	0.581	1.000	3.489	Lovo-Gómez et al. 2007
Conchagua	CON	23	9	16	0.583	0.957	5.249	Lovo-Gómez et al. 2007
Nueva Concepcion	COP	9	9	7	0.583	0.944	5.250	Lovo-Gómez et al. 2007
Izalco	IZA	11	9	11	0.701	1.000	6.309	Lovo-Gómez et al. 2007
Panchimalco	PAN	11	9	9	0.570	0.964	5.127	Lovo-Gómez et al. 2007
Nicaragua	NIC	161	9	106	0.624	0.981	5.612	Nuñez et al. 2010
Average					0.510	0.915	3.913	

4.2.3. Among-population Variation

4.2.3.1. AMOVA

Two hierarchical models were tested using AMOVA, one grouped according to major geographic region and the other according to major language group. Only loci missing less than 5% of the data were included, leaving a total of six loci used in these analyses (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS393). The results of the first AMOVA are presented in Table 4.11, where populations were grouped by major geographic area. While there is population structure, the majority of the

geographical areas. This is shown in the low percentage of variation among groups, and the statistically significant high level of diversity among populations. Additionally, the fixation indices reveal the same pattern, with high ϕ_{ST} (0.9048) and ϕ_{SC} (0.90654) and low ϕ_{CT} (-0.01890), with the low ϕ_{CT} explaining the degree of fixation due to the covariance of haplotypes among groups.

Table 4. 11. Results for the AMOVA analysis in which populations are grouped by major geographical region. Statistically significant results ($p<0.05$) are marked with an asterisk.

Source of Variation	**D of f**	**SSD**	**σ^2**	**% of σ^2**
Among groups	1	3979.856	-0.9958	-1.89
Among Populations within groups	22	32436.977	48.66039*	92.37
Within Populations	818	4103.459	5.01645*	9.52
Total	841	40520.292	52.68104	

The results of the AMOVA with populations clustered into major language group are given in Table 4.12. Like with the AMOVA of geography, little variation is explained by the covariance of haplotypes among language groups. The majority of the variation can be explained by the differences among populations within each group. This is illustrated by the high ϕ_{SC} (0.9162), giving the degree of structure explained among populations with groups, and the low ϕ_{CT} (0.13097). Finally, ϕ_{ST} was equal to 0.9272, which reveals differences among populations among groups.

Table 4. 12. Results of the AMOVA analysis in which populations were grouped into major language group. Statistically significant results are marked with an asterisk ($p<0.05$).

Source of Variation	**D of f**	**SSD**	**σ^2**	**% of σ^2**
Among groups	3	9683.757	9.92984	13.1
Among Populations within groups	18	23694.196	60.36443*	79.62
Within Populations	531	2931.871	5.52142*	7.28
Total	552	36309.825	75.81568	

4.2.3.2 Median Joining Network Analysis

The result of the MJ Network Analysis for Native American Haplogroup Q in the Poqomchi and Ch'orti' are presented in Figure 4.13. A reduced median network was used for the Y STRs to remove some of the reticulations and view the skeletal structure of the network. DYS389I and DYS389II were removed from the calculation as these are actually two markers that represent four stretched of sequence repeats. These markers need to be sequenced and

included as separate markers to avoid miscalculation. While this network reveals the diversity within haplogroup Q, there is no distinct star-like structure surrounding a large node (indicative of population expansion), satellite nodes (indicative of population substructure), or long isolate branches (indicative of genetic drift).

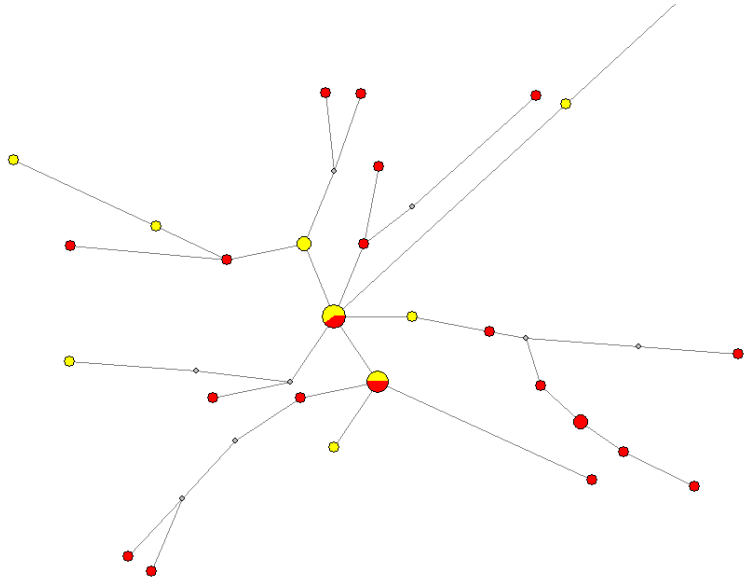

Figure 4. 13. Median-Joining Network analysis of haplogroup Q for both the Ch'orti Maya and the Poqomchi' Maya. The nodes are proportionate to the frequency of that haplotype in the population. The Poqomchi' are highlighted in red while the Ch'orti' are in yellow. The Poqomchi' sample contains one individual whose node lies outside the range of the figure.

4.2.3.3 Multi-dimensional Scaling Plots

Both monotonic and linear multi-dimensional scaling plots were generated for two and three dimensions. A two dimensional (2D) linear MDS is shown in Figure 4.14, which resulted in the lowest stress value (0.08537), indicating a good fit of the plot to the original distance matrix. The group clustered on the left of the plot (populations Guaymi-Oso, Huetar, Cabecar, Bribri, and Teribe) possesses low haplotypic diversity. While they each have significant

admixture, all non-native haplogroups are unknown. The remaining populations cluster in the upper right quadrant of the plot. Within this group, Ch'orti and Poqomchi' Maya from Guatemala cluster together, and all of the populations from Nicaragua form a cluster with the Honduran population. The populations from Panama and Costa Rica are dispersed throughout the plot.

4.2.3.4 Neighbor- Joining Tree

A neighbor-joining tree (NJT) constructed using the Kimura 2-p distances is shown in Figure 4.15. A cophenetic distance matrix was generated from the NJT and a mantel test performed comparing the cophenetic matrix to the original distance matrix. The mantel test revealed a statistically significant and high correlation between the two distance matrices (r = 0.97675, p=0.001). This indicates that the NJT is a good representation of the relationship among the populations. Like in the MDS plots, the Guaymi-Oso, Huetar, Cabecar, Bribri, and Teribe cluster together on a single branch (see node 1). Again, all of the other populations cluster together on another branch (see node 2) with the populations from El Salvador and Honduras clustering closer together. However, the Ch'orti' and Poqomchi' are not on a single branch for the NJT, despite their close relationship displayed in the MDS. Instead, the Ch'orti' split off early in the hierarchy and the Poqomchi' are clustered together with the other more distant populations from the upper right quadrant of the MDS (Kuna, Zapaton Huetar, and Ngobe).

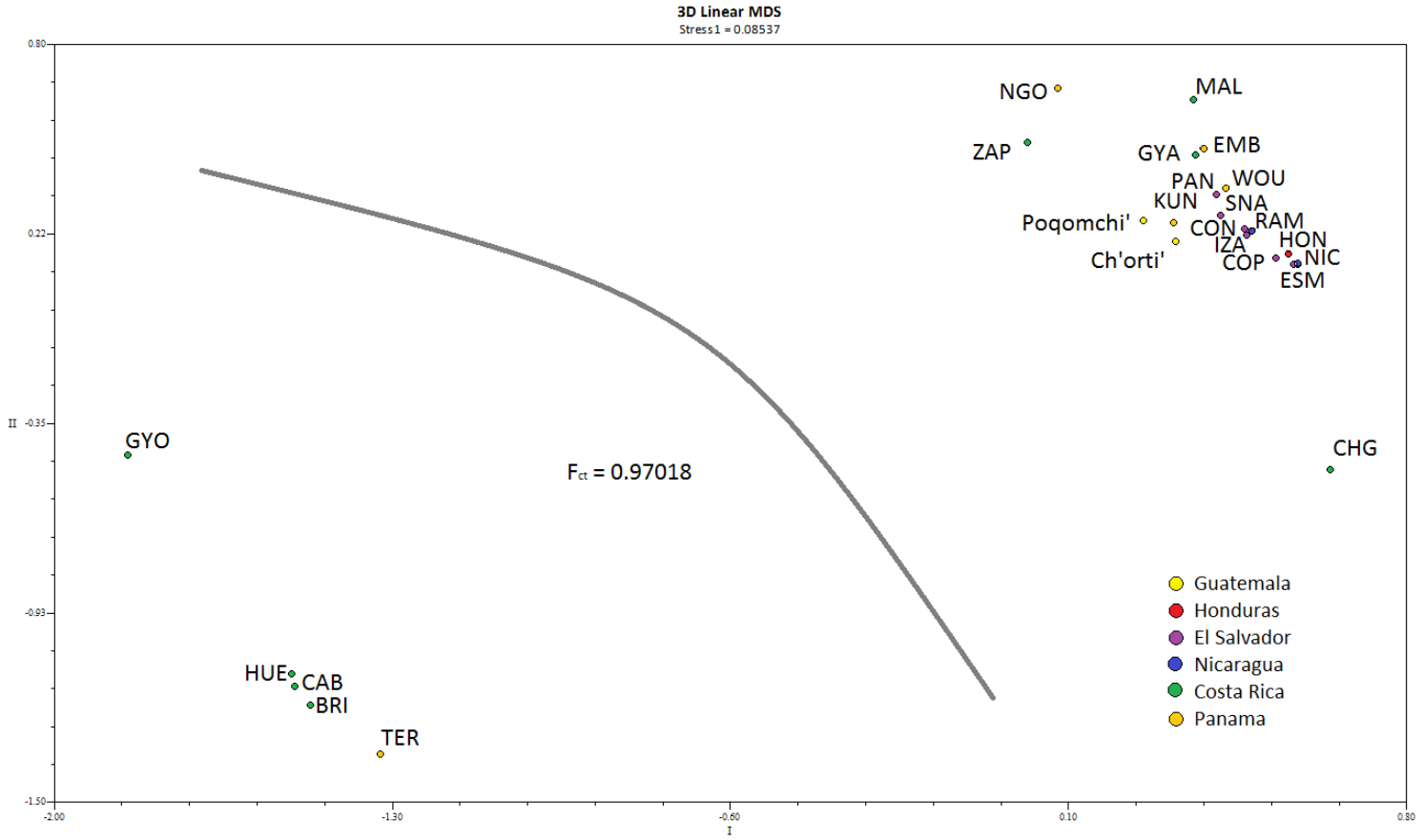

Figure 4. 14. Two-dimensional MDS plot ($p<0.05$) along with the genetic barrier detected with SAMOVA that presents the highest F_{CT} score.

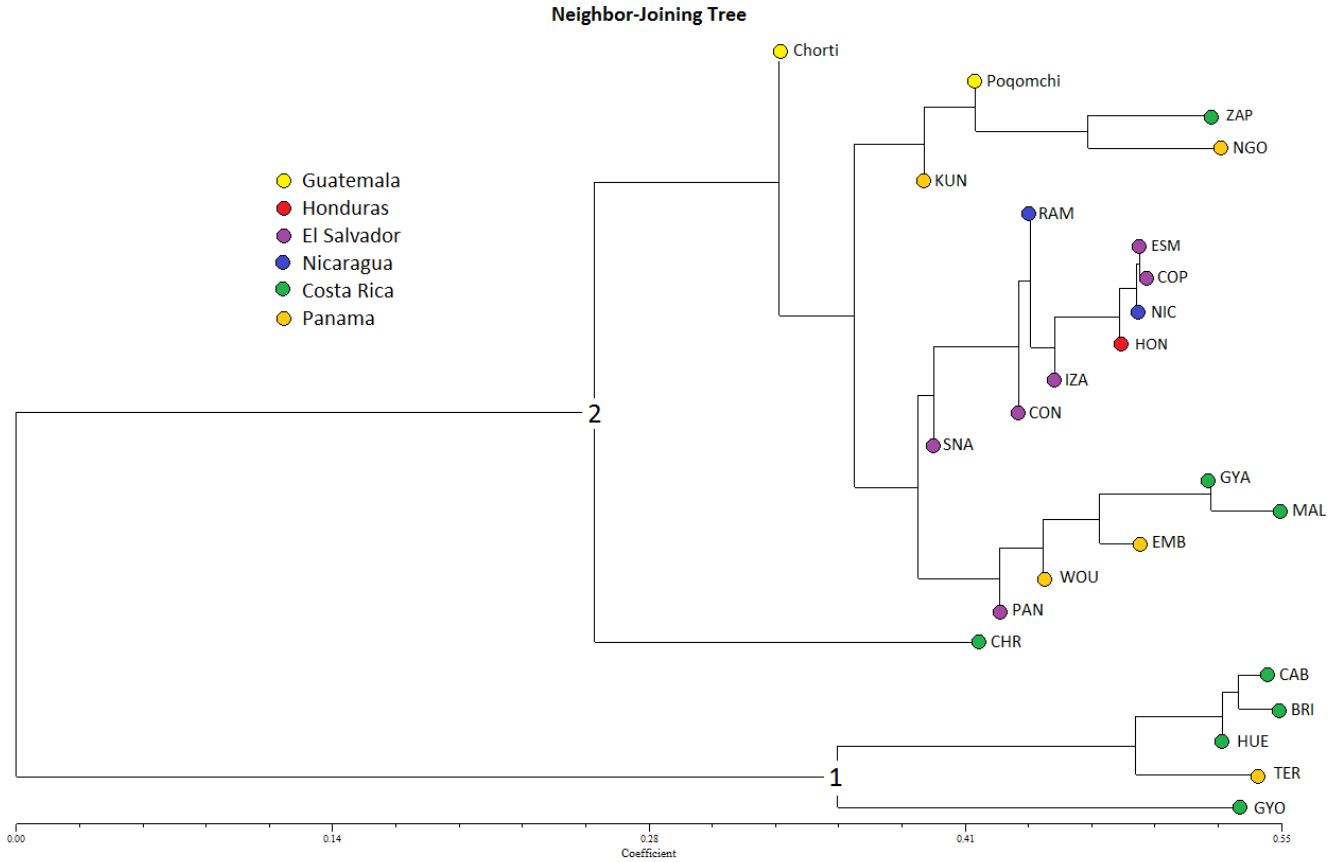

Figure 4. 15. The NJT displays the hierarchical clustering of all of the comparative populations using Slatkin's linearized R_{ST}.

4.2.4 Forces of Evolution

4.2.4.1. Mismatch Analysis

The results of the Mismatch analysis for pairwise repeat differences for the Y STR data are displayed as histograms in Figure 4.16. Only haplogroup Q (characteristic of Native American populations) was used for the analysis, as the inclusion of multiple haplogroups would inflate the probability of multiple peaks. For the Ch'orti Maya, the MPD is 3.051, and the histogram is bimodal with a peak at one repeat and five repeat differences. While the first peak is not at zero, this may be due to the small sample size and therefore the histogram may still be indicative of population stability. Alternatively, this peak may indicate a past population expansion coupled with subsequent genetic drift or selection. The Poqomchi' MPD is 4.759, with a unimodal distribution for the frequency of pairwise repeat differences, indicating a population expansion.

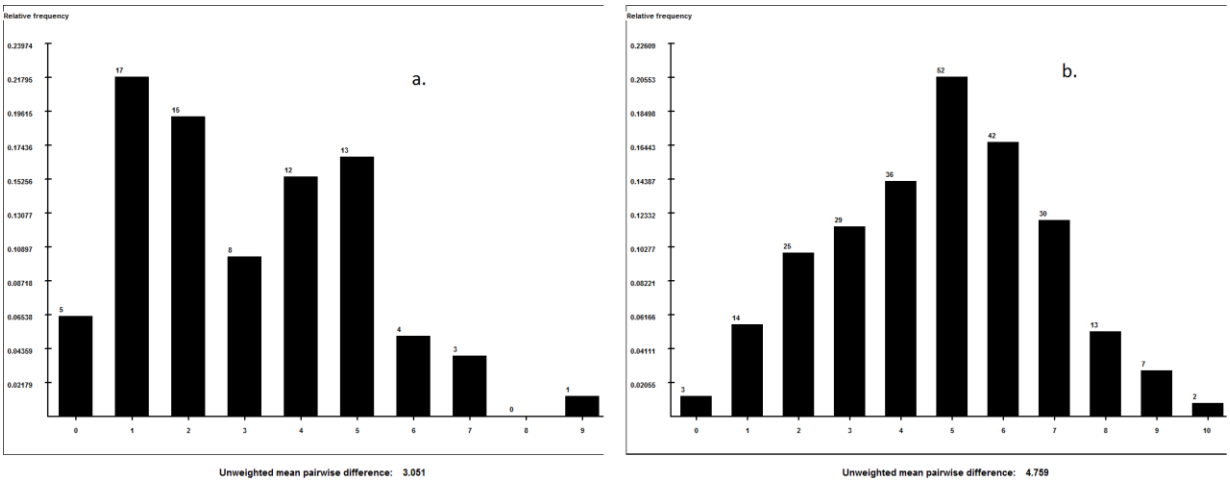

Figure 4. 16. Each histogram represents the frequency and proportionate frequency of pairwise differences for Haplogroup Q among individuals in the Ch'orti' Maya (a) and the Poqomchi' Maya (b).

4.2.4.2 Diversity vs. r_{ii}

The results of the regression of gene diversity (*h*) on distance from the centroid (r_{ii}) of haplogroup frequency differences is given in Figure 4.17. There is a weak and positive relationship between the two measure of variation (r^2= -0.0276) indicating that the assumption of the isolation by distance model does not hold true for these populations. In this instance, the Ch'orti' and Poqomchi' Maya lie above the regression line, exhibiting more variation than expected due to their European paternal admixture and population expansion in the Y Poqomchi' lineages.

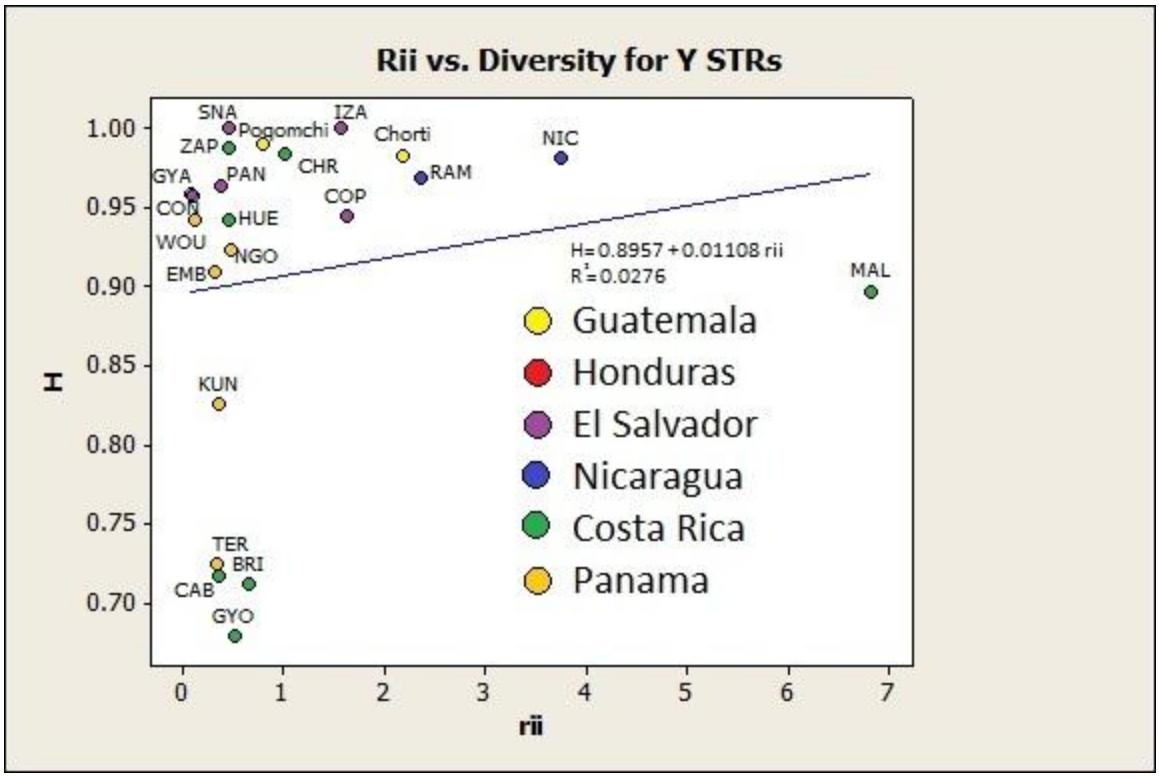

Figure 4. 17. Scatter plot and associated regression line for the relationship between gene diversity and r_{ii} for the comparative populations (plotted in Minitab v 14).

4.2.5 Phylogeography

4.2.5.1 Mantel Tests

A Mantel Randomization test comparing geographic distance and genetic distance resulted in a non-significant and inconsequential correlation (r = -0.08464, p = 0.1822). The comparison of linguistic distance and genetic distance also resulted in a very small and non-significant correlation (r = 0.01361, p = 0.5520). Like with mtDNA, there is no direct relationship between geography to genetics or language to genetics.

4.2.5.2 SAMOVA

SAMOVA was used to identify those populations that were geographically adjacent and could form groups to maximize among group variance. After running SAMOVA searching for two to five groups, the highest F_{ct} value resulted from choosing two groups of populations. These groups matched those clusters identified in the MDS plot and NJT. Figure 4.14 illustrates the barrier suggested by SAMOVA with an F_{ct} of 0.9718 ($p<0.001$). Again, the group clustered on the bottom left of the plot possesses low haplotypic diversity and includes the Guayami-Oso, Huetar, Cabecar, Bribri, and Teribe. While they each have significant admixture, all non-native haplogroups are unknown.

4.2.5.3 Monmonier's Maximum Difference Algorithm

Monmonier's Maximum Difference Algorithm was used to identify five possible genetic barriers among the populations used in this study (See Figure 4.18). The first barrier, highlighted in red, begins along the border of the populations separating Central American populations and MesoAmerican populations, and then encircles the Maléku population from Central America. The second barrier, highlighted in yellow, detects a barrier between the two Maya populations and all other populations. The third barrier, in green, separates the Panamanian populations, exclusive of the Teribe, from all other populations. The blue barrier is the fourth barrier detected by Monmonier's, which separates the Huetar from the other populations in the plot.

The final barrier to gene flow identified (purple) is between the Izapan and the adjacent Mesoamerican populations.

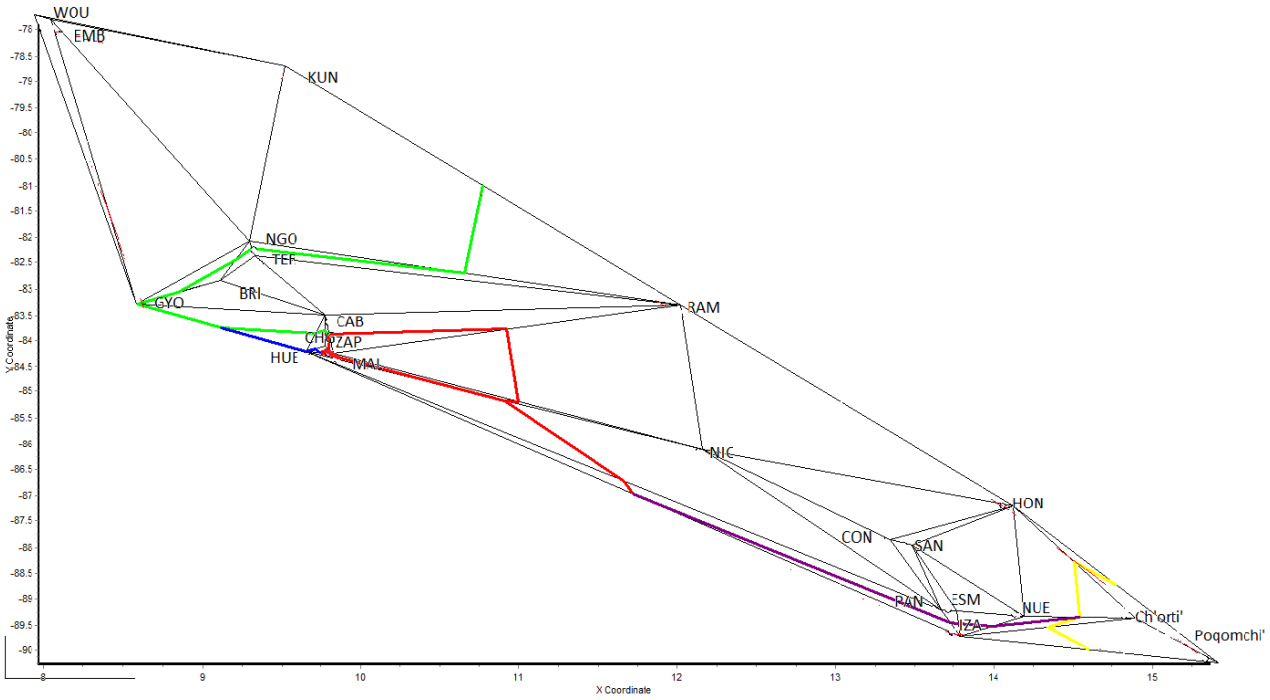

Figure 4. 18. Barriers chosen by Monmonier's Maximum Difference Algorithm. The red barrier was the first barrier detected, yellow the second, third is green, fourth is blue, and fifth barrier is purple.

4.2.5.4 Interpolated Genetic Landscape

Genetic distances were regressed on geographic distance and the resulting residuals were used to create an interpolated genetic landscape, which is shown in Figure 4.19. Blue peaks represent greater distance than expected, indicative of genetic drift, while red dips show

greater genetic similarity, indicative of gene flow. There is a clear change in the degree of population similarity between the Mesoamericans and northern Central American populations. This figure highlights the greater similarity among the Mesoamerican populations and greater divergence among the Central American populations. The sharpest nadir in the genetic landscape (greatest similarity) exists between the mixed Honduran population and the nearest El Salvadoran populations. The residual scores change between negative and positive values between the El Salvadoran and Honduran borders with Nicaragua. The highest peaks (areas of greatest genetic divergence along geographic borders) exist between the northwestern Nicaraguan populations and the other Central Americans. These patterns mimic the patterns observed in the other phylogeographic methods.

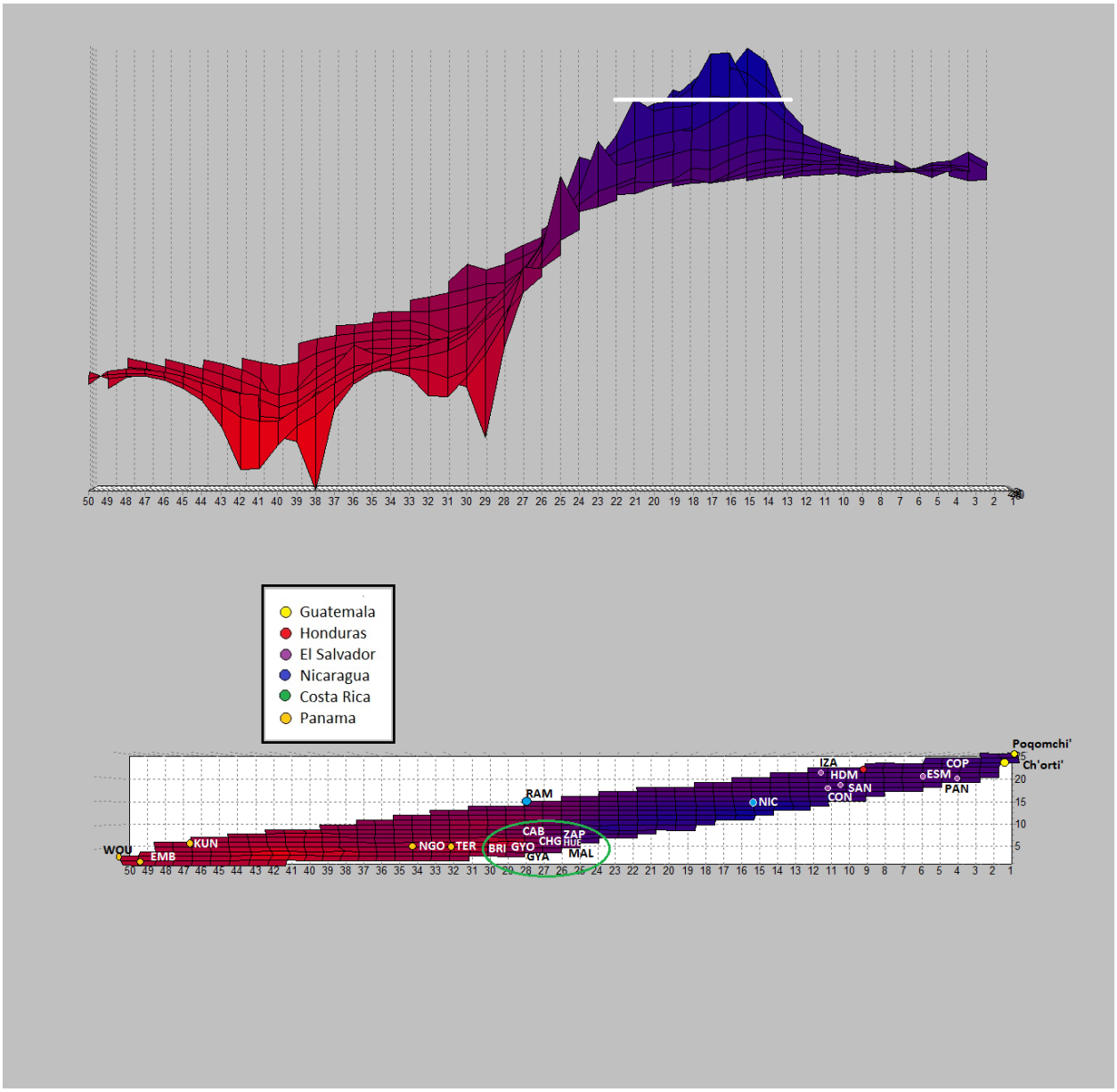

Figure 4. 19. Interpolated genetic landscape (GL) using residual genetic distances. Blue peaks represent greater distance than expected, indicative of genetic drift, while red dips show greater genetic similarity, indicative of gene flow. The white bar denotes the zero axes for the z plane. All points below this are indicative of population similarity.

CHAPTER 5. DISCUSSION

5.1 INTRODUCTION

The following chapter combines the results presented in the previous chapter with findings in genetics, language and history to make broad observations about the hypotheses under investigation. The discussion will first break down the various types of analyses conducted, including presence of major mtDNA and Y chromosome lineages and within group variation. This is followed by discussions on variation among populations, evidence of evolution operating on these populations, and phylogeography of the Americas. Finally, this chapter will integrate these observations with a discussion of the hypotheses proposed in chapter one, including: 1) presence of genetic structure among Maya-speaking populations; 2) variation among Maya and other Meso-, Central, and South American and Caribbean populations; 3) differing patterns of maternal (inferred from mtDNA) and paternal genetic structure (inferred from the Y-chromosome); 4) the statistical relationship between geography, genetics and languages across Meso-, Central, and South America and the Caribbean.

5.2 LINEAGES AND WITHIN GROUP VARIATION

Not surprising, the mtDNA lineages present in the Poqomchi' and Ch'orti' are consistent with previous findings in the Mesoamerican cultural and geographical region. For most of Mesoamerica, all four major Native American mtDNA haplogroups (A, B, C, and D) are present

(Sandoval et al. 2009; Schurr 2004). However, the Poqomchi' Maya only exhibit A, B, and C; and for the Ch'orti' only A and C haplogroups are found. It is more common among the Central American populations to have missing haplogroups (e.g. only the Wounan exhibit all four haplogroups in appreciable frequencies) (Kolman and Bermingham 1997). So, the absence of these other haplogroups in the Poqomchi' and Ch'orti' may be indicative of genetic drift, poor sampling, or a close relationship with nearby Central American neighbors as a result of gene flow. Also, for both the Poqomchi' and Ch'orti' Maya populations, haplogroup A has the highest frequency. This is the case for all of the Mesoamerican populations except for the Tarahumara from northern Mexico, who exhibit a high frequency of haplogroup C. There is much more diversity in haplogroup frequencies across Central and South America than in Mesoamerica. In the Central American Isthmus, the populations have high frequencies of haplogroup A, but Central America is interspersed with populations with high frequencies of haplogroup B (e.g., Emberrá and Rama) and haplogroup C (Wounan). Overall, the frequency of haplogroup A is lower in Central America (Kolman and Bermingham 1997; Melton 2008). In South America, the western highlands are characterized by high frequencies of haplogroup B, while eastern and southern South America vary greatly by population regarding more frequent haplogroups (Bert et al. 2004; Fuselli et al. 2003; Ginther et al. 1993; Justice et al. In Press; Lewis et al. 2005; Melton 2008; Melton et al. 2007; Merriwether et al. 2000; Rickards et al. 1999; Schmitt et al. 2004; Schurr and Sherry 2004; Ward et al. 1996; Williams et al. 2002). Both ancient Caribbean populations studied to date have high frequencies of haplogroup C (Lalueza-Fox et al. 2001; Lalueza-Fox et al. 2003).

The most common mtDNA haplotype among the Poqomchi' Maya is A2 (16111T, 16223T, 16290T, 16319A, 16362C), which is present in appreciable frequency among all Mesoamerican populations except for the Tarahumara (who are not part of the Mesoamerican cultural region, but reside in Mexico) and the K'iche Maya (Boles et al. 1995; Sandoval et al. 2009; Torroni et al. 1993a). This haplotype is also present in considerable frequency among most Central American populations, but is less frequent among South American populations. This haplotype is also the most common within the Yucatec Maya, and in high frequency within the Ch'orti' Maya, representing a deep lineage within Mesoamerica, that may have been subsequently lost in the K'iche during one of the episodes of population reduction (e.g., European Colonization, Guatemalan Civil War). The most common haplotype for the Ch'orti Maya is a population specific haplotype within C1 (16111T, 16223T, 16244A, 16274A, 16298C, 16325C, 16327T). It is also important to note that none of the individuals with this haplotype reported having the same mother, although two individuals failed to report their mother's name. All nine individuals were residents of different aldeas (hamlets) within the municipalities of Jocotán or Camotán. The absence of this common Ch'orti' haplogroup among the other Mayan populations and the absence of the A2 founding lineage in the K'iche Maya is indicative of heterogeneity among the populations.

Alternatively, the haplogroup C cluster in the Ch'orti' Maya may be a remnant of the Mesoamerican peripheral groups with whom the Ch'orti' Maya may have been admixed. It is hypothesized that the Ch'orti' Maya were the prominent ruling class in Copan, but that the peasant class may have been largely non-Maya and most likely Lenca in origin (Metz et al. 2009). In fact, previous investigations of ancient remains in Copan, have revealed a difference

between the modern Maya, with high mtDNA haplogroup A, and the past population, with the absence of haplogroup A and high frequency of haplogroup C (Gonzalez-Oliver et al. 2001). While these results are not likely representative of the greater Copan population due to small sample size (n=9), they may be indicative of a unique genetic structure in pre-contact eastern Guatemala. The city of Jocotán was founded by forced relocation of Indians from the Copán region; so, one explanation for the unique and common haplogroup C lineages within the Ch'orti' is that they may be a remnant from the peasant class of Copán (Grünberg and Misión de Verificación de las Naciones Unidas en Guatemala. 2003; Reina 1969).

For mtDNA, all Maya populations have above average gene and nucleotide diversity, with the highest gene diversity among the Maya in the Poqomchi' (h = 0.947) and highest nucleotide diversity in the Ch'orti (π = 0.020). These values are even higher than some known admixed populations (e.g., El Salvador) (Salas et al. 2009) and populations that are known to have undergone demographic and spatial expansion (Nahua) (Sandoval et al. 2009), indicating the relatively high level of diversity within these populations despite lacking one or more of the Native American mtDNA haplogroups. Both populations that exhibit the lowest gene and nucleotide diversities, the Ijka and Ache, are near fixation for one haplogroup (A and B respectively).

When compared to mtDNA, Y chromosome lineages indicate greater heterogeneity in the study populations. For both the Poqomchi' and Ch'orti', haplogroup Q is the most common with the majority also belonging to the Native American specific lineage of Q1a3a1-M3. Neither population exhibited Native American haplogroup C-M130 or C3b-P39, which are both found among Native American populations in North America. The Ch'orti' exhibit evidence of non-

Native admixture indicative of European gene flow (haplogroups I, J, E1b1a), and the Poqomchi' have one unidentified non-native haplogroup. European admixture is present on the paternal side in all populations used for comparison; however, the Poqomchi' display the lowest proportion of non-Native lineages. The pooled Nicaraguan population exhibits the highest degree of admixture, but not the highest of any of the three diversity measurements. The Ch'orti' and Poqomchi', despite little admixture, both display above average values for haplotype and MPD diversity measures. These results suggest that the native lineages remaining in Meso- and Central America are heterogeneous.

The Poqomchi', K'iche, and Yucatec Maya exhibit only mtDNA haplogroups A, B, C, and D indicating the absence of non-Native American admixture. However, three individuals in the Ch'orti' region possess non-Native haplogroups (two have haplogroup L2 and one unidentified European). All three individuals reported Spanish as their primary language and ethnicity, but did not report language and ethnicity for their mothers. Both individuals who exhibit L2 have different mothers and different haplotype motifs. Furthermore, the paternal lineages reveal more significant admixture in the Ch'orti', including European haplogroups I and J, and haplogroup E which is present in both the Mediterranean and eastern Africa. In contrast, the Poqomchi' only possesses one paternal lineage that may be due to non-native gene flow.

While the Poqomchi' were peacefully converted after colonization (Cahuec del Valle 1997), the Ch'orti region underwent a great population reduction due to not only diseases such as smallpox, but slavery, starvation, and warfare. Jocotán, itself, was directly adjacent to a trade route, which drew Spanish and Ladinos of mixed descent in from the capital, especially during the height of the indigo trade 1600-1800 AD. During this time the population continued to

experience repeated plagues, famine, and near slave-like conditions on indigo, cacao, tobacco, sugarcane, cattle, and other plantations. To maintain the labor force, African slaves were brought in to assist in transporting goods along the trade route as well as work the fields on the various plantations (Lovell and Lutz 1995; Metz 2006). The presence of European and African lineages (mtDNA: L2 and unidentified; Y: I, J, and E) within this population and the deep structured lineages (highlighted by the network analysis) are a testament to the impact of European colonization on the genetic makeup of the Ch'orti' Maya.

5.3 VARIATION AMONG POPULATIONS

Previous studies have shown conflicting results regarding genetic variation among Mayan populations. Using common Alu insertions, Herrera et al. 2007 found that a significant amount of heterogeneity exists among the Mayan populations (Kakchikel, K'iche, and Yucatec). However, another investigation on Mayan genetic variation concluded that the Mayan are among the most homogenous cultural groups in the Americas (Ibarra-Rivera et al. 2007). This study used autosomal STRs to examine the genetic makeup on the Choles, Yucatec, K'iche, Kakchikel, and Huastec) and found that the Huastec were the only significantly different Mayans. Similarly, this study has shown that little substructure exists among the maternal ancestry of the Maya included here (Yucatec, Ch'orti', K'iche, and Poqomchi'). This is evidenced by the negative ϕ_{CT} score and the lack of any variation among groups in the mtDNA AMOVA

analysis. Also, the mtDNA network analysis reveals that there are many haplotypes shared among the populations, and satellite nodes are often only one mutational step apart.

Additionally, the most divergent Mayan population for the mtDNA markers is difficult to identify given that each test implies a different Mayan population. For example, the NJT displays the greatest branch length difference between the Poqomchi' and the other three populations, while the MDS shows that the K'iche do not cluster closely with the others, and the network analysis exhibits satellite clusters more common within the Ch'orti' and K'iche populations. All comparisons between Poqomchi' and Ch'orti' Maya using Y chromosome markers indicate a close relationship between the two. The inconsistency among these tests is a further demonstration of the close genetic relationship among the Mayans, as a strong underlying genetic structure should be clearly discernible in each test.

Significant population structure is observed outside of the Maya region. The presence of maternal population substructure is shown in the moderate fixation indices and statistically significant proportion of variation explained among major geographic (ϕ_{CT} = 0.12049) and linguistic (ϕ_{CT} = 0.10739) groups in the mtDNA AMOVA analysis. Interestingly, these results are replicated in the Y chromosome for major linguistic groups (ϕ_{CT} = 0.13097), but not for geographical group (ϕ_{CT} = -0.01890), but this difference could be due to the lower number of populations available for comparison or differential gene flow with Europeans and Africans.

Despite the high proportion of variation that exists among major geographic regions for mtDNA, the MDS plots do not reveal a strong geographic pattern. The Maya populations tend to cluster toward the center of the plot, but do not cluster as tightly together as one would expect given the AMOVA results. The Poqomchi' are the most central population, likely due to

their high frequency of A2, which is shared across most populations in the study. This would also explain why the K'iche Maya are the furthest from the Poqomchi' of the other Mayan populations, as they lack A2. The majority of the Central American populations are located along the right side of the plot with the Andean and western Amazonian populations; however, the other South American populations and the Mexican populations tend to be more dispersed throughout the plot. While this study has used a different method of estimating genetic distance for constructing the MDS plot, the results mimic other studies which show the existence of a close relationship between Central America and western South America, greater variation in Amazonia, and a close relationship between Caribbean populations. Differences between this study and previous ones can be found in the placement of the plot's peripheral populations (e.g., Tarahumara, Otomi) (Lewis et al. 2007; Melton 2008; Melton et al. 2007; Sandoval et al. 2009). The Y chromosome MDS plots exhibit a similar pattern, with the Poqomchi' and the Ch'orti' clustering closely together, but with the Central American populations being more dispersed. The complicated relationship among these Native American populations is reflected in the NJTs for mtDNA and Y chromosome data. In both cases, the trees display no clear geographic or historical picture and are a poor fit to the original genetic distance matrix.

5.4 EVOLUTIONARY FORCES AND NEUTRALITY

Information gathered from the neutrality test statistics, network analyses, mismatch plots, diversity *versus* r_{ii} plots, and genetic interpolation all indicate that forces of evolution and demographic changes are affecting the genetic makeup of the populations under study. Both the Ch'orti' and Poqomchi' Maya, as seen in the r_{ii} *versus* diversity plot, possess above average diversity. This variation is a result of a combination of factors, including maintaining a long-standing genetic relationship to surrounding Mesoamericans, non-Native gene flow, and having undergone recent population expansions following a deep bottle neck. The Poqomchi' and the Ch'orti' exhibit negative scores for both neutrality test statistics, and both were statistically significant for Fu's F_s, which is sensitive to recent population expansion. Population expansion is further supported in the network analyses, where an excess of singletons is noticeable for both populations, and in the unimodal mismatch analyses for haplogroup A. The maintenance and expansion of this diversity has not been maintained in isolation. Evidence that this expansion occurred after a population bottleneck rests in the network analyses. The network analysis for haplogroup A reveals a star-like structure, especially when the Maya samples are pooled, which indicate that the population is expanding. The mismatch plots of the pooled Maya for haplogroup A also reveal a unimodal distribution and a nearly unimodal plot for haplogroup B. Overall, the Maya exhibit an expanding population. However, when the Poqomchi' and Ch'orti' mtDNA haplogroup A networks are separated, one can see that there are deep and fragmented lineages (especially with the Ch'orti') indicative of past population reduction, which caused the

loss of intermediate lineages. A population bottleneck is even more obvious in the Y chromosome haplogroup Q network analysis and mismatch analysis. In both instances, it seems the Ch'orti' have been more affected by genetic drift then has the Poqomchi'. Historical estimates of population size for the Poqomchi' and Ch'orti' indicate that the Ch'orti' were affected more by European colonization than were the Poqomchi' Additionally, the Poqomchi' have undergone a more significant population expansion since the end of the Guatemalan Civil War (Figure 2.8).

In addition to a recent population expansion, gene flow has played a role in the level of diversity present in the Poqomchi' and Ch'orti' Maya. Both populations are above the theoretical regression line in the r_{ii} *versus* diversity plots for mtDNA, indicating gene flow into these populations. To a large extent, this gene flow is occurring with other Mayan populations, but may also involve other nearby Mesoamerican indigenous populations. It has already been demonstrated that the Poqomchi' and Ch'orti' share a close relationship with the Yucatec and the K'iche, but further analyses reveal close ties to surrounding Mesoamericans as well. All Mesoamerican populations exhibit negative Tajima's D scores and either negative or near zero scores for Fu's F_s and lie above the regression line in the r_{ii} *versus* diversity plot for mtDNA. Therefore, these populations are unaffected by drift and are undergoing expansion and/or exhibit higher than expected diversity due to gene flow. However, all of the Central Americans except for the Wounan, exhibit positive scores for these measures, indicating that they are experiencing drift. The pattern of gene flow across Mesoamerica and drift across Central America is illustrated in the interpolated genetic landscape for both the mtDNA sequence data and Y chromosome STRs.

5.5 PHYLOGEOGRAPHY

Phylogeographic methods reveal an interesting picture of genetic relationships across geographic and linguistic space. For both mtDNA and Y chromosome, there are no correlations between genetics and geography or linguistics using the Mantel randomization of the distance matrices. This would seem to indicate that there is no relationship between these factors/features. However, the results of the AMOVA and SAMOVA for mtDNA reveal that there is a relationship between geography and language with genetics, but that this relationship cannot be expressed in a pairwise correlation, but instead in a comparison of major geographic and linguistic groups.

Mesoamerica is a geographic region that stretches across lower Mexico, Guatemala, Belize, Honduras, and El Salvador, but is defined not only by geography, but shared cultural characteristics. As part of the culture, language is shared. It is hypothesized that the linguistic characteristics that tie these groups together are a result of continued contact through trade, the use of a common trade language, and a shared cosmological view (Campbell 1997; Carmack et al. 2007; Coe 2005). The connection between geography, culture, and language has created a similar grouping of populations for the AMOVA analyses. However, it is more striking that using the SAMOVA, which does not include predefined groups, identifies with few exceptions almost identical groups when trying to maximize the amount of variation among groups. The first group identified by SAMOVA included almost all of the Mesoamerican populations except the Tarahumara and Otomi, which have been shown to be the most divergent of the Mexican populations (this study [see MDS, Figure 4.4], (Sandoval et al. 2009). Campbell (1997) has

suggested that the Tarahumara are outside of the Mesoamerican cultural and linguistic group and thus fits nicely with their exclusion in this analysis. Also included in the Mesoamerican group for the SAMOVA are the most divergent of the Central and South American populations with high frequencies of haplogroup A (Ijka, Kogi, Cayapa, Guatuso Maleku, Guayami, Ngobe, and the mixed Nicaragua sample). Many of these populations are identified as outliers in their geographic location by the Monmonier's maximum difference algorithm (e.g., Otomi, Ngobe, Guateuso Maleku). Finally, the most striking evidence of a strong relationship between genetics and geography is found in both the mtDNA and Y chromosome interpolated genetic landscapes. In both cases, it is evident that there is significant admixture among the Mesoamerican populations and genetic drift among the Central American populations. This pattern is less evidence in the Y chromosome data, due to the great similarity among the Chibchan populations from Costa Rica and Panama. These populations (Cabecar, Bribri, Triqui, Huetar, and Guaymí) were selected based on previous evidence that the participants were of primarily Native American descent. These samples were part of a previous study where classical markers were used to remove those individuals that showed Euroepan admixture (Ruiz-Narvaez et al. 2005). This sampling bias has created an artificial landscape of heightened similarity for the Y markers among the Chibchan-speaking group. Despite this difficulty, it remains apparent that the Mesoamerican groups are closely related to one another and more distant to the Central American populations [See Figure 4.19]. The South American populations exhibit gene flow in the west and genetic drift in the east for mtDNA sequence data.

5.6 HYPOTHESES

5.6.1 Genetic Structure of the Mayans

This study set out to determine if the Mayan populations of Mesoamerica should be considered a biologically homogenous population distinct from surrounding Latin American populations. As evidenced by remains of large game animals and simple unifacially-edged and retouched tools, archaeologists know that hunter-gatherers resided in Mesoamerica and in what we refer to now as the Maya regions as early as 20,000 BC (Stark 1981). However, these populations grew in size and began exploiting the environment leading to a subsequent reduction in diversity of exploited food resources including fauna and plants. In turn this led to the domestication of plants evidenced by the increase in domesticated varieties of maize, avocado, chilies, beans, and squash. These developments along with the invention of vessels that could be used for long-term storage allowed populations to settle down during the Archaic Period (~8000 - 4000 BC) (Stark 1981). Archaeological and linguistic evidence indicates that it is in this period when the shared common ancestors of all Mayan-speaking populations settled down and expanded into the Guatemalan highlands where the modern day Kaqchikel still live (Campbell 1997; Carmack et al. 2007; Coe 2005; Vogt 1969). Following continued demographic growth, the population fissioned and the Huastec, followed by the Yucatec-Lakandon, then Ch'olan (Western language branch), and finally the Mamean language group, split from the parental population spreading throughout modern Mexico, Belize, Guatemala, El Salvador, and Honduras. While these populations diverged linguistically, material remains, hieroglyphics, and

shared cultural views held these populations together through time and helped to foster communication and trade. This study provides evidence that these material exchanges also offered opportunities to maintain close biological relationships through the pattern of dominant maternal and paternal lineages, as seen in the mtDNA and Y chromosome MJ Networks, and SAMOVA, Monmonier Maximum Difference Algorithm, and interpolated genetic landscape, and the mtDNA AMOVA. In each instance it was shown that the Mayan population under study (Ch'orti', Poqomchi', Yucatec, and K'iche) have experienced gene flow with nearby populations and only minimal variance exists between these populations.

This is not to say that there is no population substructure among Mayan populations, but little measureable structure. There are indications that the Maya cannot be considered a biologically homogenous population distinct from surrounding populations. The mtDNA ANOVA comparing the major Maya geographic regions indicates that 6.55% of the variation can be explained by differences among populations within groups. However, it should be noted that due to the small number of Mayan populations for which mtDNA data are available, only the Poqomchi' and K'iche Maya are grouped into the southern Maya region, and only the Yucatec represent the northern and Ch'orti' the central Maya area. Therefore, the majority of this variation is due to the differences between the Poqomchi' Maya and the K'iche Maya. If we are to assume that these populations are representative of their regions, this difference can be explained. Previously, the K'iche have been shown to be the most divergent of the Mayan populations using autosomal STRs (Ibarra-Rivera et al. 2007) They diverge from the other Maya in this study by their absence of the mtDNA A2 haplotype and a higher frequency of haplogroup D sequences that are not shared with the other populations. Furthermore, previous genetic

studies using polymorphic *Alu* insertions (Herrera et al. 2007) and ancient ondontometrics (Scherer 2007) have shown differences between the southern Maya area highland populations and other Maya areas. It is likely that the movement of the Poqomchi' Maya into the eastern highlands increased their cultural and biological contact with the other Mayans.

5.6.2 Genetic Structure of the Americas

The genetic relationship between the Maya and surrounding Mesoamerican populations further reveals that the Maya are not an entirely homogenous group distinct from surrounding populations. As mentioned previously, Mesoamerica is distinguishable as a cultural and geographic region based upon shared archaeological history and a number of distinct linguistic traits based on shared linguistic features such as phonology, syntax, and word and sentence morphology. Mesoamerica is inhabited by populations that are assumed to represent language groups descended from a common ancestor prior to the Formative Period (Campbell 1997; Carmack et al. 2007; Grove 1981). Similar diets, shared domesticated plant varieties, use of the same calendar, common trade languages, and shared iconography are further indications of continued and extensive contact throughout Mesoamerica. This strong geographic and cultural connection has also led to an apparent genetic relationship as well. Mesoamerican is connected by shared lineages, especially high frequencies of haplogroups A and A2, and little population substructuring, as evidenced in the AMOVAs for language and geography, and the SAMOVA. Several tests reveal that the similarities among populations are maintained through gene flow.

These tests include the position of Mesoamerican populations above the regression line in the r_{ii} *versus* diversity plots, negative residual distances in the Interpolated Genetic Landscape, and high diversity measurements. The continued gene flow among populations makes it difficult to determine phylogenetic relationships resulting in inconsistent outcomes between the NJT and MDS.

Overall, there seems to be a genetic barrier between the southeastern border of Mesoamerica and Central America. This study and previous studies all indicate that the populations of Central America, dominated by Chibchan speaking populations, diverged some 6,000 BP, maintaining only minimal contact (Melton et al. 2007). Isolation here resulted in the "tribalization" of these populations such as those populations of the Amazon (Schurr 2004). These phenomena are implied by the mtDNA and Y chromosome SAMOVA, positive mtDNA neutrality tests scores, position of most Central American populations below the regression line in the mtDNA r_{ii} *versus* diversity plot, and the population dispersion in the mtDNA and Y chromosome MDS plots. While Melton (2008) found a close relationship between Chibchan populations in Central America and the K'iche, the present study does not replicate these findings, likely due to the inclusion of a greater number of comparative Mesoamerican and Mayan populations.

5.6.3 Maternal and Paternal Genetic Variation

Differences between maternal and paternal lineages exist for the Maya populations under study. The obvious difference between mtDNA and Y chromosome is in the effect of non-native admixture. After European colonization, the Native American populations declined drastically. The Spanish required a labor force, but many native men died from warfare, disease and if they survived the initial conquest, they often died from mistreatment in slavery. As a result, African slaves were transported to Guatemala as a supplementary work force. The great majority of the slaves were men, and in the absence of African women, they often sought partners among the Indians of Guatemala. Additionally, the Spanish colonizers were also predominantly male (Lovell and Lutz 1995). The effects of the male admixture were not equal across all of Guatemala. Since the Dominicans peacefully converted the Poqomchi' and surrounding highland population of the Verapaz, many more men survived in this area and families were able to stay on their land. Due to their greater survival rates, slaves were not needed as a labor force in this region. Also, Europeans and Ladinos were more likely to move to towns with important commodities and located along trade routes. The Alta Verapaz had two main staples, coffee and sugar, which were not important until the late 1800s. In contrast, the Ch'orti region was of great import for Ladinos due to its location along a trade route, indigo, sugar, and other cash crops. As mentioned above, the Ch'orti' were also more affected by population decline creating a need for supplementary labor force. These historical differences resulted in a sharp contrast in the proportion of admixture between the Poqomchi' and the

Ch'orti' and between male and female lineages. The Poqomchi' do not exhibit any non-native gene flow on the maternal side; however, one non-native paternal lineage is evident (4%). For the Ch'orti' non-native gene flow is present for both mtDNA and Y chromosome, but more significant male admixture is shown (26.3% for male and 5.3% for female).

Maternal and paternal lineages reflect the same pattern of gene flow and genetic drift. Also, greater heterogeneity and signatures of genetic drift are evident among the Central American populations *versus* the Mesoamericans for both mtDNA and Y chromosomes. Even without the comparative populations from South America, it is clear that the Y chromosome STRs show greater gene flow in Mesoamerica when compared to Central America, as evident by the interpolated genetic landscape, SAMOVA, and Monmonier's maximum difference algorithm. Finally, both maternal and paternal investigations show a similar picture of demographic expansion in the Poqomchi' Maya through unimodal mismatch analyses and a high frequency of unique haplotypes for the Native American haplogroups. However, while mtDNA haplogroup A in the Ch'orti' exhibit population expansion, Y chromosome haplogroup Q indicates a greater effect of genetic drift. As discussed above, this difference is to be expected for the male lineages in the Ch'orti' region given the greater population reduction at European contact, the introduction of non-Native male lineages, and evidence of a slower post-contact recovery in population size.

5.6.4 Relationship Between Geography, Language, and Genetics

As mentioned above in the discussion of phylogeographic methods, a strong relationship between geography, language, and genetics exists for Native American populations. However, the results of the AMOVA, SAMOVA, and interpolated genetic landscape for mtDNA reveal that there is a relationship between geography and language with genetics, but that this relationship cannot be expressed in a pairwise correlation of distances, but instead in a comparison of major geographic and linguistic groups. The Y chromosome reveals that a statistically significant proportion of variation can be explained through language, but not geography in the Mantel randomization tests.

CHAPTER 6. CONCLUSIONS

This project explored the genetic makeup of the Poqomchi' Maya and Ch'orti' Maya and compared their genetic structure with the greater Latin American geographic and cultural regions. Both maternal and paternal lineages investigated in this study revealed a strong correlation with both the historical and linguistic record. The mtDNA lineages present in the Poqomchi' and Ch'orti' are consistent with previous findings in the Mesoamerican cultural and geographical region, but also reflect the unique histories of both these populations.

While previous studies have painted conflicting pictures regarding variation among Mayan populations and their affiliation with outside groups (Gonzalez-Oliver et al. 2001; Herrera et al. 2007; Ibarra-Rivera et al. 2008; Melton 2008; Scherer 2007) this study has shown consistently that the Mayan populations share a common history and a close genetic relationship. The mtDNA AMOVA analysis for the Maya exhibited little population substructuring. Also, the mtDNA network analysis reveals that there are many haplotypes shared among the populations and satellite nodes are infrequent and usually one mutational step apart. This study supports previous findings that the K'iche Maya of the southern Maya area are the most divergent of the Mayan groups (Ibarra-Rivera et al. 2008), and diverge from the other Maya in this study by their absence of the mtDNA A2 haplotype and the higher frequency of haplogroup D. Furthermore, previous genetic studies using polymorphic *Alu* insertions (Herrera et al. 2007) and ancient ondontometrics (Scherer 2007) have shown differences between the southern Maya area highland populations and other Maya areas, but these differences seem to have affected the Poqomchi' Maya to a lesser degree.

The Maya also fit well within the variation of the greater Mesoamerican cultural and linguistic landscape, which can be considered biologically distinct from the rest of Latin America. This is evidenced for both mtDNA and Y chromosome data in the AMOVA analyses and each of the phylogeographic methods. For both the Poqomchi' and Ch'orti' Maya populations, mtDNA haplogroup A and Y haplogroup Q had the highest frequency, and mtDNA sequence and Y STR haplotypes within these haplogroups are shared with populations throughout Mesoamerica. There is greater diversity in haplogroup frequencies across Central and South America indicative of their unique pattern of southward peopling from Central America across the Andes, then western migration followed by isolation and tribalization. The presence of maternal population substructuring is shown in the moderate fixation indices and statistically significant proportion of variation explained among major geographic (ϕ_{CT} = 0.12049) and linguistic (ϕ_{CT} = 0.10739) groups in the mtDNA AMOVA analysis, and in the Y chromosome for major linguistic groups (ϕ_{CT} = 0.13097). These results are likely weighted by by one study. The participants of the Chibchan populations from Costa Rica (Bribri, Teribe, Cabecar, Huetar, Guayami) were all selected to minimize non-native admixture by removing individuals with European blood group markers (Ruiz-Narvaez et al. 2005). This sampling procedure would artificially make these populations, who share a common language family, more similar genetically. The Mesoamerican populations share a common ancestral population, as do all Native Americans, but the unique similarities among Mesoamerican populations are maintained through gene flow. Unfortunately, the continued gene flow between populations

makes it difficult to determine phylogenetic relationships resulting in inconsistent outcomes between the NJT and MDS.

Despite nearly 500 years of colonization, the Poqomchi' and Ch'orti' Maya maintain a majority of Native American mtDNA and Y chromosome lineages. Differences between maternal and paternal lineages exist for the Maya populations under study, but vary by geographic region. Since the Dominicans peacefully converted the Poqomchi', population reduction was not as severe as in other regions of Guatemala. Both men and women survived colonization and highland Guatemala saw a rapid demographic recovery during the post-colonization era. Furthermore, the Alta Vera Paz did not become commercially important until the late 1800s and therefore populations like the Poqomchi' remained biologically relatively isolated from the Spanish, Ladino, and African populations (Carmack 1986; Feldman 2000). As a result, non-native admixture is rare within this population exhibited by only one non-native paternal lineage. Both parental markers suggest a similar picture of demographic expansion in the Poqomchi' Maya through unimodal mismatch analyses, high levels of diversity, and a high frequency of unique haplotypes for the Native American haplogroups.

In contrast, the Ch'orti region was of great import for Ladinos due to its location along a trade route, and the presence of indigo, sugar, tobacco, and other cash crops. As mentioned above, the Ch'orti' were also more affected by population decline creating a need for supplementary labor force. These historical differences have resulted in a sharp contrast in the proportion of admixture between the Poqomchi' and the Ch'orti' and between male and female lineages. For the Ch'orti' non-native gene flow is present for both mtDNA and Y chromosome, but more significant male admixture is shown (26.3% for male and 5.3% for female).

The Ch'orti Maya along with most of eastern, northern, and southern Guatemala underwent a great population reduction due to not only diseases such as smallpox, but slavery, starvation, and warfare. Jocotán itself was directly adjacent to the major trade route and was prominent in the commercial production of indigo, cacao, tobacco, and minerals. The Spanish and Ladino populations, along with their African slaves, were thus drawn from the city to this region, particularly during the height of the indigo trade (1600-1800 AD). As a result, the Ch'orti' exhibit European and African gene flow on both the maternal and paternal side, but with more significant admixture on the paternal side, including European haplogroups I and J, and haplogroup E, which is present in both the Mediterranean and eastern Africa. While new genes were introduced through gene flow, conflicts over land, poor working conditions, and continued epidemics continued to restrict the gene pool more in the Ch'orti' region than in highland Guatemala. This impact was greatest on the paternal side, as the Y chromosome haplogroup Q indicates a greater effect of genetic drift. The presence of European and African lineages within this population and the deep structured lineages highlighted by the network analysis, are testament to the impact of European colonization on the genetic makeup of the Ch'orti' Maya.

The results of this study provide evidence of close genetic relationship among Mayan populations and to their neighboring Mesoamerican populations. This shared genetic relationship supports previous findings of a greater continuity of biological relationships along geographic distance as compared to linguistics; however, this relationship is due to strong cultural ties, not only geographic proximity. While there are differences between the maternal and paternal lineages, these differences are not as marked as in previous studies on the genetic

structure of Native American populations (Rubicz et al. 2010; Zlojutro et al. 2009). Once again, molecular markers have proven useful in helping elucidate the historical, geographic, and linguistic relationship among recently diverged human populations.

Bibliography

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, and Bandelt HJ. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. PLoS One 3(3):e1764.

Avise JC. 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology 7(4):371-379.

Avise JC. 2000. Phylogeography : the history and formation of species. Cambridge, Mass. ; London: Harvard University Press. viii, 447 p. p.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CE, and Saunders C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18:489-522.

Bandelt HJ, Forster P, and Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16(1):37-48.

Bandelt HJ, Forster P, Sykes BC, and Richards MB. 1995. Mitochondrial portraits of human populations using median networks. Genetics 141(2):743-753.

Batista O, Kolman CJ, and Bermingham E. 1995. Mitochondrial DNA diversity in the Kuna Amerinds of Panama. Hum Mol Genet 4(5):921-929.

Bert F, Corella A, Gene M, Perez-Perez A, and Turbon D. 2004. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. Ann Hum Biol 31(1):9-28.

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, and Lopez-Camelo JS. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63(6):1862-1871.

Boles TC, Snow CC, and Stover E. 1995. Forensic DNA testing on skeletal remains from mass graves: a pilot project in Guatemala. J Forensic Sci 40(3):349-355.

Bonatto SL, Redd AJ, Salzano FM, and Stoneking M. 1996. Lack of ancient Polynesian-Amerindian contact. Am J Hum Genet 59(1):253-258.

Bonatto SL, and Salzano FM. 1997a. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am J Hum Genet 61(6):1413-1423.

Bonatto SL, and Salzano FM. 1997b. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci U S A 94(5):1866-1871.

Bortolini MC, Salzano FM, Bau CH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-De-Guerra D, Bedoya G et al. . 2002. Y-chromosome biallelic polymorphisms and Native American population structure. Ann Hum Genet 66(Pt 4):255-259.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT et al. . 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73(3):524-539.

Cahuec del Valle E, editor. 1997. Historia y Memorias de la Comunidad Etnica Poqomchi': Universidad Rafael Landivar.

Campbell L. 1997. American Indian languages : the historical linguistics of Native America. New York: Oxford University Press. xi, 512 p. p.

Carmack RM. 1986. Ethnohistory of the Guatemalan Colonial Indian. In: Spores R, editor. Ethnohistory. Austin, TX: University of Texas Press.

Carmack RM, Gasco J, and Gossen GH. 2007. The legacy of Mesoamerica : history and culture of a Native American civilization. Upper Saddle River, N.J.: Pearson/Prentice Hall. xviii, 558 p. p.

Carmack RS. 1988. Harvest of violence : the Maya Indians and the Guatemalan crisis. Norman: University of Oklahoma Press. xvii, 334 p. p.

Coe MD. 1966. The Maya. London,: Thames & Hudson. 252 p. p.

Coe MD. 2005. The Maya. New York: Thames and Hudson. 272 p. p.

Cooke R. 2005. Prehistory of Native Americans on the Central American Land Bridge: Colonization, Dispersal and Divergence. Journal of Archaeological Research 13(2):129-187.

Crawford MH. 1998. The origins of Native Americans : evidence from anthropological genetics. Cambridge ; New York: Cambridge University Press. xv, 308 p. p.

de Azevedo S, Nocera A, Paschetta C, Castillo L, Gonzalez M, and Gonzalez-Jose R. 2011. Evaluating microevolutionary models for the early settlement of the New World: The importance of recurrent gene flow with Asia. Am J Phys Anthropol 146(4):539-552.

Dietz G. 2005. From Indigenismo to Zapatismo. In: Postero NG, and Zamosc L, editors. The Struggle for Indigenous Rights in Latin America. Portland, Oregon: Sussex Academic Press.

Dupanloup I, Schneider S, and Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11(12):2571-2581.

Ersts PJ. 2011. Geographic Distance Matrix Generator. 1.2.3 ed: Center for Biodiversity and Conservation. p available from http://biodiversityinformatics.amnh.org/gdmg/comments.php.

Evans ST, and Webster DL. 2001. Archaeology of ancient Mexico and Central America : an encyclopedia. New York: Garland Pub. xliii, 948 p. p.

Excoffier L, and Lischer HE. 2010a. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564-567.

Excoffier L, and Lischer HE. 2010b. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567.

Excoffier L, Smouse PE, and Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131(2):479-491.

Excoffier L, and Yang Z. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16(10):1357-1368.

Faulhaber J. 1970. Anthropometry of Living Indians. In: Stewart TD, editor. Physical Anthropology. Austin, Texas: University of Texas Press.

Feldman LH. 2000. Lost shores, forgotten peoples : Spanish explorations of the South East Mayan lowlands. Durham N.C.: Duke University Press. xxiv, 269 p. p.

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, and Brinkmann B. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67(1):182-196.

Foxen P. 2007. In search of providence : transnational Mayan identities. Nashville, Tenn.: Vanderbilt University Press. xxiv, 358 p. p.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147(2):915-925.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, and Pettener D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. Mol Biol Evol 20(10):1682-1691.

Gabbert W. 2004a. Becoming Maya : ethnicity and social inequality in Yucatan since 1500. Tucson: University of Arizona Press.

Gabbert W. 2004b. Of Friends and Foes: The Caste War and Ethnicity in Yucatan. The Journal of Latin American Anthropology 9(1):90-118.

Geppert M, Baeta M, Nunez C, Martinez-Jarreta B, Zweynert S, Cruz OW, Gonzalez-Andrade F, Gonzalez-Solorzano J, Nagy M, and Roewer L. 2011. Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. Forensic Sci Int Genet 5(2):100-104.

Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, Just J, Salzano FM, and King MC. 1993. Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. EXS 67:211-219.

Gonzalez-Oliver A, Marquez-Morfin L, Jimenez JC, and Torre-Blanco A. 2001. Founding Amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo. Am J Phys Anthropol 116(3):230-235.

Grimes BF. 1992. Ethnologue Index. Dallas, TX: Summer Institute of Linguistics.

Grove DC. 1981. The formative period and the evolution of complex culture. In: Brickner VR, editor. Archaeology. Austin, TX: University of Texas Press.

Grünberg G, and Misión de Verificación de las Naciones Unidas en Guatemala. 2003. Tierras y territorios indígenas en Guatemala. Guatemala: Facultad Latinoamericana de Ciencias Sociales (FLACSO-Guatemala) : Misión de Verificación de las Naciones Unidas (MINUGUA) : Dependencia Presidencial de Asistencia Legal y Resolución de Conflictos sobre la Tierra (CONTIERRA). 166 p. p.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, and Zegura SL. 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15(4):427-441.

Hammer MF, and Zegura SL. 2002. The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. Annual Review of Anthropology 31:303-321.

Harpending H, and Ward BE. 1982. Chemical systematics and human populations. In: Nitecki M, editor. Biological Aspects of Evolutionary Biology. Chicago, IL: University of Chicago Press.

Herrera RJ, Rojas DP, and Terreros MC. 2007. Polymorphic Alu insertions among Mayan populations. J Hum Genet 52(2):129-142.

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, and Tajima K. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol Biol Evol 10(1):23-47.

Ibarra-Rivera L, Mirabal S, Regueiro MM, and Herrera RJ. 2008. Delineating genetic relationships among the Maya. Am J Phys Anthropol 135(3):329-347.

Jobling M, Hurles M, and Tyler-Smith C. 2004. Human evolutionary genetics : origins, peoples & disease. New York: Garland Science. xx, 523 p. p.

Jones GD. 1989. Maya resistance to Spanish rule : time and history on a colonial frontier. Albuquerque: University of New Mexico Press. xvii, 365 p. p.

Justice A. 2007. Apportionment of Genetic Variation in Contemporary Aleut and Eskimo Populations of Alaska Using Anthropometrics and Classical Genetic Markers. Lawrence, Kansas: University of Kansas.

Justice A, Dean B, and Crawford MH. In Press. Molecular Consequences of migration and urbanization in the Peruvian Amazonia. In: Crawford MH, and Campbell B, editors. Causes and Consequences of Human Migration. New York, NY: Cambridge University Press.

Justice A, Rubicz R, Chittoor G, Jantz RL, and Crawford MH. 2010. Anthropometric variation among Bering Sea natives. Hum Biol 82(5-6):653-675.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, and Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18(5):830-838.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P et al. . 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet 64(3):817-831.

Kaufman T. 1976. Archeological and Linguistic Correlations in Mayaland and Associated Areas of Meso-America. World Archaeol 8(1):101-118.

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA et al. . 2004. A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74(6):1183-1197.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111-120.

Kolman CJ, and Bermingham E. 1997. Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. Genetics 147(3):1289-1302.

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, and Guionneau-Sinclair F. 1995. Reduced mtDNA diversity in the Ngobe Amerinds of Panama. Genetics 140(1):275-283.

Lalueza-Fox C, Calderon FL, Calafell F, Morera B, and Bertranpetit J. 2001. MtDNA from extinct Tainos and the peopling of the Caribbean. Ann Hum Genet 65(Pt 2):137-151.

Lalueza-Fox C, Gilbert MT, Martinez-Fuentes AJ, Calafell F, and Bertranpetit J. 2003. Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. Am J Phys Anthropol 121(2):97-108.

Lauria-Santiago A, and Binford L. 2004. Landscapes of struggle : politics, society, and community in El Salvador. Pittsburgh, Pa.: University of Pittsburgh Press. ix, 326 p. p.

Lewis CM, Jr., Lizarraga B, Tito RY, Lopez PW, Iannacone GC, Medina A, Martinez R, Polo SI, De La Cruz AF, Caceres AM et al. . 2007. Mitochondrial DNA and the peopling of South America. Hum Biol 79(2):159-178.

Lewis CM, Jr., Tito RY, Lizarraga B, and Stone AC. 2005. Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. Am J Phys Anthropol 127(3):351-360.

Lovell WG, and Lutz C. 1995. Demography and empire : a guide to the population history of Spanish Central America, 1500-1821. Boulder, Colo. ;: Oxford : Westview Press. xv, 190 p. p.

Lutz C, Carmack RM, and Early JD. 1982. The Historical demography of highland Guatemala. Albany, N.Y.: Institute for Mesoamerican Studies, State University of New York at Albany. v, 202 p. p.

Malhi RS, Gonzalez-Oliver A, Schroeder KB, Kemp BM, Greenberg JA, Dobrowski SZ, Smith DG, Resendez A, Karafet T, Hammer M et al. . 2008. Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history. Am J Phys Anthropol 137(4):412-424.

Manahan TK. 2008. Anatomy of a Post-Collapse Society: Identity and Interaction in Early Postclassic Copán. In: Stanton TW, and Magnoni A, editors. Ruins of the past : the use and perception of abandoned structures in the Maya lowlands. Boulder: University Press of Colorado. p xviii, 364 p.

Manly BFJ. 2005. Multivariate statistical methods : a primer. Boca Raton, FL: Chapman & Hall/CRC Press. 214 p. p.

Manni F, Guerard E, and Heyer E. 2004. Geographic patterns of (genetic, morphological, linguistic) variation: How barriers can be detected using Monmonier's algorithm. Human Biology 76(2):173-190.

Martínez Novo C. 2006. Who defines indigenous? : identities, development, intellectuals, and the state in northern Mexico. New Brunswick, N.J.: Rutgers University Press. xi, 187 p. p.

Mateus Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM, and Herrera RJ. 2005. Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. Am J Phys Anthropol 128(1):171-184.

Matson GA. 1970. Distribution of Blood Groups. In: Stewart TD, editor. Physical Anthropology. Austin, Texas: University of Texas Press.

Melton P. 2008. Genetic history and pre-Columbian Diaspora of Chibchan speaking populations: Molecular genetic evidence. Lawrence, KS: University of Kansas. 231 p.

Melton PE, Briceno I, Gomez A, Devor EJ, Bernal JE, and Crawford MH. 2007. Biological relationship between Central and South American Chibchan speaking populations: evidence from mtDNA. Am J Phys Anthropol 133(1):753-770.

Merriweather DA, and Kaestle FA. 1999. Mitochondrial recombination? (continued). Science 285(5429):837.

Merriwether DA, Kemp BM, Crews DE, and Neel JV. 2000. Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA. In: Renfrew C, editor. America Past, America Present: Genes and Languages in the Americas and Beyond. Cambridge, UK: Cambridge University Press.

Merriwether DA, Rothhammer F, and Ferrell RE. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol 98(4):411-430.

Metz B. 2009. The Ch'orti' Area. In: Metz B, McNeil CL, and Hull KM, editors. The Ch'orti' Area: Past and Present. Gainesville, FL: University Press of Florida.

Metz BE. 2006. Ch'orti'-Maya survival in eastern Guatemala : indigeneity in transition. Albuquerque: University of New Mexico Press. ix, 346 p. p.

Metz BE, McNeil CL, and Hull KM. 2009. The Ch'orti' Maya area : past and present. Gainesville: University Press of Florida. 346 p. p.

Meyer S, Weiss G, and von Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152(3):1103-1110.

Miller MP. 2005. Alleles in Space: computer software for the joint analysis of Interindividual spatial and genetic information. Journal of Heredity 96:722-724.

Monmonier MS. 1973. Maximum-Difference Barriers: An Alternative Regionalization Numerical Method. Geographical Analysis 5(3):245-261.

Nei M. 1987. Molecular Evolutionary Genetics. New York, NY: Columbia University Press.

Nei M, and Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76(10):5269-5273.

Nelson TM, Just RS, Loreille O, Schanfield MS, and Podini D. 2007. Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing. Croat Med J 48(4):460-472.

Nunez C, Baeta M, Sosa C, Casalod Y, Ge J, Budowle B, and Martinez-Jarreta B. 2010. Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers. Am J Phys Anthropol 143(4):591-600.

Olson JS. 1991. The Indians of Central and South America : an ethnohistorical dictionary. New York: Greenwood Press. xiii, 515 p. p.

Pine A. 2008. Working hard, drinking hard : on violence and survival in Honduras. Berkeley, Calif. ; London: University of California Press. xiv, 253 p. p.

Ramirez-Soriano A, and Nielsen R. 2009. Correcting estimators of theta and Tajima's D for ascertainment biases caused by the single-nucleotide polymorphism discovery process. Genetics 181(2):701-710.

Reina R. 1969. Chapter 6. In: Vogt EZ, editor. Ethnology. Austin, TX: University of Texas Press. p 101-133.

Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, and Cann RL. 1999. mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. Am J Hum Genet 65(2):519-530.

Roberts DF, and Newman MT. 1970. Physiological Studies. In: Stewart TD, editor. Physical Anthropology. Austin, Texas: University of Texas Press.

Rogers AR, and Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9(3):552-569.

Rogers AR, and Jorde LB. 1996. Ascertainment bias in estimates of average heterozygosity. Am J Hum Genet 58(5):1033-1041.

Rohlf FJ. 2000. NTSYSpc. 2.1 ed: Exeter Software.

Rothhammer F, Llop E, Carvallo P, and Moraga M. 2001. Origin and evolutionary relationships of native Andean populations. High Alt Med Biol 2(2):227-233.

Rubicz R, Schurr TG, Babb PL, and Crawford MH. 2003. Mitochondrial DNA variation and the origins of the Aleuts. Hum Biol 75(6):809-835.

Rubicz R, Zlojutro M, Sun G, Spitsyn V, Deka R, Young KL, and Crawford MH. 2010. Genetic architecture of a small, recently aggregated Aleut population: Bering Island, Russia. Hum Biol 82(5-6):719-736.

Ruhlen M. 1987. A guide to the world's languages. Stanford, Calif.: Stanford University Press. v. <1 > p.

Ruiz-Narvaez EA, Santos FR, Carvalho-Silva DR, Azofeifa J, Barrantes R, and Pena SD. 2005. Genetic variation of the Y chromosome in Chibcha-speaking Amerindians of Costa Rica and Panama. Hum Biol 77(1):71-91.

Sabloff JA. 1990. The new archaeology and the ancient Maya. New York: Scientific American Library : Distributed by W.H. Freeman. xii, 193 p. p.

Saitou N, and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406-425.

Salas A, Lovo-Gomez J, Alvarez-Iglesias V, Cerezo M, Lareu MV, Macaulay V, Richards MB, and Carracedo A. 2009. Mitochondrial echoes of first settlement and genetic continuity in El Salvador. PLoS One 4(9):e6882.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, and Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74(3):454-465.

Salzano FM. 2002. Molecular variability in Amerindians: widespread but uneven information. An Acad Bras Cienc 74(2):223-263.

Sandoval K, Buentello-Malo L, Penaloza-Espinosa R, Avelino H, Salas A, Calafell F, and Comas D. 2009. Linguistic and maternal genetic diversity are not correlated in Native Mexicans. Hum Genet 126(4):521-531.

Santos M, Ward RH, and Barrantes R. 1994. mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. Hum Biol 66(6):963-977.

Scherer AK. 2007. Population structure of the Classic period Maya. Am J Phys Anthropol 132(3):367-380.

Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, and Salzano FM. 2004. Extremely limited mitochondrial DNA variability among the Ache Natives of Paraguay. Ann Hum Biol 31(1):87-94.

Schneider S, and Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152(3):1079-1089.

Schurr TG. 2004. The Peopling of The New World: Perspectives from Molecular Anthropology. Annual Review of Anthropology 33:551-583.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46(3):613-623.

Schurr TG, and Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol 16(4):420-439.

Seielstad MT, Minch E, and Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. Nat Genet 20(3):278-280.

Sharer RJ. 2009. The Ch'orti Past: An Archaeological Perspective. In: Metz B, McNeil CL, and Hull KM, editors. The Ch'orti' Area: Past and Present. Gainesville, FL: University Press of Florida.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139(1):457-462.

Smith DG, Malhi RS, Eshleman J, Lorenz JG, and Kaestle FA. 1999. Distribution of mtDNA haplogroup X among Native North Americans. Am J Phys Anthropol 110(3):271-284.

Stanton TW, and Magnoni A. 2008. Ruins of the past : the use and perception of abandoned structures in the Maya lowlands. Boulder: University Press of Colorado. xviii, 364 p. p.

Stark B. 1981. The rise of sedentary life. In: Brickner VR, editor. Archaeology. Austin, TX: University of Texas Press.

Sturrock K, and Rocha J. 2000. A multidimensional scaling stress evaluation table. Field Methods 12(1):49-60.

Sullivan PR. 2000. The Yucatec Maya. In: Reilfler Bricker V, editor. Ethnology. Austin, Texas: University of Texas Press.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105(2):437-460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585-595.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, and Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68(6):1485-1496.

Torres Moss JC. 1996. Apuntes para la historia de Jocotán. Guatemala: Editorial Cultura, Ministerio de Cultura y Deportes. 88 p. p.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, and Wallace DC. 1993a. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53(3):563-590.

Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, and Wallace DC. 1993b. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53(3):591-608.

Underhill PA, Jin L, Zemans R, Oefner PJ, and Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci U S A 93(1):196-200.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P et al. . 2000. Y chromosome sequence variation and the history of human populations. Nat Genet 26(3):358-361.

Vigilant L, Stoneking M, Harpending H, Hawkes K, and Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science 253(5027):1503-1507.

Vogt EZ. 1969. Chapter 2. In: Vogt EZ, editor. Ethnology. Austin, TX: University of Texas Press. p 21-30.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C et al. . 2007. Genetic variation and population structure in native Americans. PLoS Genet 3(11):e185.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, and Santos RV. 1996. Mitochondrial DNA polymorphism in three Brazilian Indian Tribes. American Journal of Human Biology 8:317-323.

Watanabe JM. 2000. Maya and Anthropologists in the Highlands of Guatemala since the 1960s. In: Reilfler Bricker V, editor. Ethnology. Austin, Texas: University of Texas Press.

Watanabe JM. 2001. Maya and Anthropologists in the Highlands of Guatemala since the 1960's In: Monaghan JD, editor. Supplement to the Handbook of Middle American Indians. Austin, YX: University of Texas Press.

Williams SR, Chagnon NA, and Spielman RS. 2002. Nuclear and mitochondrial genetic variation in the Yanomamo: a test case for ancient DNA studies of prehistoric populations. Am J Phys Anthropol 117(3):246-259.

Wright S. 1951. The genetical structure of populations. Annual Eugenics 15:323-354.

Wright S. 1965. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. Evolution 19(3):395-420.

Zegura SL, Karafet TM, Zhivotovsky LA, and Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21(1):164-175.

Zlojutro M, Rubicz R, and Crawford MH. 2009. Mitochondrial DNA and Y-chromosome variation in five eastern Aleut communities: evidence for genetic substructure in the Aleut population. Ann Hum Biol 36(5):511-526.

APPENDICES

APPENDIX A

The Department of Anthropology at the University of Kansas supports the practice of protection for human subjects participating in research. The following information is provided for you to decide whether you wish to participate in the present study. You should be aware that even if you agree to participate, you are free to withdraw at any time without penalty.

We are conducting this study to better understand anemia, nutrition, and genetic makeup of the population in the municipality of Tamahú, Guatemala. This will entail your completion of an oral interview. The interview is expected to take approximately 15 minutes to complete. Also, a doctor or nurse will draw 3 milliliters of blood to test for anemia. If you do have anemia the results will tell us what type of anemia you have and how severe. The results of your test will be presented to you within one week. The health center can help you recuperate from anemia, provide medication, and explain what foods strengthen your blood. Prior to the blood withdraw will be asked to give a saliva sample for assistance in the DNA portion of the study.

The content of the interview should cause no more discomfort than you would experience in your everyday life. Although participation may not benefit you directly, we believe that the information obtained from this study will help us gain a better understanding of knowledge, perceptions, and causes related to anemia. Your participation is solicited, although strictly voluntary. Your names will only be associated in this study to construct a family tree of the population. If you would like additional information concerning this study before or after it is completed, please feel free to contact us by phone or mail.

Completion of the survey indicates your willingness to participate in this project and that you are considered an adult by you community. For you time and participation we will compensate you with 3 pounds of beans and/or 3 pounds of rice, and/or 3 pounds of corn, and/or vegetable oil. If you have any additional questions about your rights as a research participant, you may call (785) 864-7429 or write the Human Subjects Committee Lawrence Campus (HSCL), University of Kansas, 2385 Irving Hill Road, Lawrence, Kansas 66045-7563, email dhann@ku.edu.

Sincerely,

James W. Herynk "Santiago"
Principal Investigator
Department of Anthropology
Fraser Hall
University of Kansas
Lawrence, KS 66045
(785) 864-4103
herynk@ku.edu

Dr. John Janzen, Ph.D.
Faculty Supervisor
Department of Anthropology
Fraser Hall
University of Kansas
Lawrence, KS 66045
(785) 864-4103
jjanzen@ku.edu

Approved by the Human Subjects Committee University of Kansas, Lawrence Campus (HSCL). Approval expires one year from 7/19/2007.

APPENDIX B

Forma oral del consentimiento

El Departamento de la Antropología en la Universidad de Kansas apoya la práctica de la protección para los sujetos humanos que participan en esta investigación. La información siguiente se le proporciona para que usted decida participar o no en el actual estudio. De esta manera usted deberá saber que aunque decida participar, usted estará libre de retirarse en cualquier momento si lo desea y sin ninguna pena o compromiso.

Estamos conduciendo este estudio para reconstruir los orígenes y las migraciones entre los indígenas americanos en Centro y Suramérica, usando la información de la genética molecular. Para tal fin, es necesario que usted llene el cuestionario y que nos proporcione una muestra de saliva de su boca y mejilla. El cuestionario tardará aproximadamente 30 minutos. La muestra biológica será utilizada para extraer el DNA. La misma solo se utilizará para reconstruir la historia genética de Centro y Suramérica. Solamente el personal que trabaja directamente en este proyecto tendrá acceso a la DNA y al cuestionario.

El contenido de los cuestionarios no tendrá ninguna repercusión sobre su vida cotidiana y no le causará ningún malestar posterior. Aunque su participación pueda no beneficiarle directamente, creemos que la información obtenida de este estudio nos ayudará a obtener una mejor comprensión de la evolución y la historia de la población de la Centro y de Suramérica, y para lo cual, solicitamos su participación y le reiteramos que su decisión es enteramente voluntaria.

Su nombre no será asociado de ninguna manera a los resultados de la investigación. Si usted quisiera la información adicional referente a este estudio, antes o después de que se termine, siéntase por favor libre de contactarnos, ya sea por teléfono o correo electrónico.

Una vez concluido el cuestionario procederemos a tomar pruebas bucales y de esputo, indicación sobre su voluntad de participar en este proyecto, y que usted es un mayor de edad por encima de los dieciocho años. Si usted tiene preguntas adicionales sobre las sus derechos como participante de la investigación, nos puede llamar al teléfono (785) 864-7429 o (785) 864-7385 o escribir al comité de temas humanos en el campus universitario en la ciudad de Lawrence:
Human Subjects Committee Lawrence Campus (HSCL), University of Kansas, 2385 Irving Hill Road, Lawrence, Kansas 66045-7563, email jbutin@ku.edu or mdenning@ku.edu.

Aprobado por Comité de Lawrence de los sujetos humanos (HSCL) en 6/5/2008. Aprobación de HSCL expira un año a partir de la 8/2/2008. HSCL#16735

Approved by the Human Subjects Committee Lawrence (HSCL) on 6/5/2008. HSCL approval expires one year from 8/2/2008. HSCL#16735

Sinceramente,

Anne Justice, M.A.
Investigador principal
Departamento de la Antropología
Jayhawk 1415 Blvd.
622 Fraser Pasillo
Universidad de Kansas
Lawrence, KS 66045, los E.E.U.U.
(785) 864-2606
ajustice@ku.edu

Michael Crawford, Ph.D.
Profesor tutor del proyecto
Departamento de la antropología
Jayhawk 1415 Blvd.
622 Fraser Pasillo
Universidad de Kansas
Lawrence, KS 66045, los E.E.U.U.
(785) 864-4170
crawford@ku.edu